



Knowledge Organization and Cultural Diversity

Organizers: José Augusto Chaves Guimarães e Vera Dodebei



ISKO-BRASIL

Organizers

José Augusto Chaves Guimarães
Vera Dodebei

| 2

Knowledge Organization and Cultural Diversity

Marília, São Paulo, Brasil

Sociedade Brasileira de Organização do Conhecimento
(ISKO-Brasil)

2017

K73k Knowledge Organization and Cultural Diversity [recurso eletrônico] / José Augusto Chaves Guimarães, Vera Dodebei, organizadores. -- Pernambuco: ISKO-Brasil ; UFPE, 2017.
760 f. ; 30 cm.

ISBN: 978-85-415-0924-4

Livro digital

1. Organização do conhecimento. I. Título.

CDD 025.4

Scientific Committee

Vera Lúcia D. Louzada de Mattos Dodebei (UNIRIO)

Gercina Ângela Borém de Oliveira Lima (UFMG)

Marisa Bräscher Basílio Medeiros (UFSC)

Maria Aparecida Moura (UFMG)

Fabio Assis Pinho (UFPE)

| 4

Editors

Isadora Victorino Evangelista

Gilberto Gomes Cândido

Rafael Aparecido Moron Semidão

Rafael Cacciolari Dalessandro

Suellen Oliveira Milani

Design

Maíra Fernandes Alencar

Translation

Natália Nakano

© Reproduction of this book in whole or in part is permitted provided that the claims mentioned.
The sale is prohibited.

SUMMARY

THE EPISTEMOLOGIC DIMENSION OF KNOWLEDGE ORGANIZATION

Approaches and paradigms in knowledge organization	30
<i>Leila Cristina Weiss e Marisa Bräscher (UFSC)</i>	
Historical-epistemological dimension in Knowledge Organization: Gesner's taxonomy contributions, 16th century	37
<i>Andre Vieira de Freitas Araujo e Giulia Crippa (USP)</i>	
Terminological and interdisciplinary research configurations in knowledge organization (2010-2014)	47
<i>Leilah Santiago Bufrem, Aline Elis Arboit e Juliana Lazzarotto Freitas (UNESP)</i>	
The conceptual dimension of Knowledge Organization in NASKO conferences: Bardian's content analysis	55
<i>José Augusto Chaves Guimarães (UNESP); Rodrigo de Sales (UFF), André Ynada dos Santos (UNESP) e Daniela Fernanda de Oliveira Matos (UNESP)</i>	
The relation between knowledge organization and information science in the Brazilian scientific community: an investigation within ISKO-Brazil	73
<i>Rodrigo de Sales (UFF)</i>	
Poole, index and fractures: indexing and serial publications in the United States of the nineteenth century	85
<i>Gustavo Silva Saldanha (IBICT/Unirio) e Naira Christofolletti Silveira (Unirio)</i>	
Processes of archival knowledge representation: historical and conceptual elements for classification and description	94
<i>Natália Bolfarini Tognoli (UNESP) e Thiago Henrique Bragato Barros (UFPA)</i>	
Ontologies and Simple Knowledge Organization Systems (SKOS): similarities and differences	101
<i>Rogério Aparecido Sá Ramalho (UFSCAR)</i>	
Academic genealogy as an approach to domain analysis	108
<i>Renata Cristina Gutierrez Castanha e Maria Cláudia Cabrini Grácio (UNESP)</i>	
The epistemological dimension of Document Analysis of fictional content in Knowledge Organization	116
<i>Roberta Caroline Vesú Alves e João Batista Ernesto de Moraes (UNESP)</i>	
Indexing languages and domain analysis	122
<i>Vera Lucia Ribeiro Guim e Mariângela Spotti Lopes Fujita (UNESP)</i>	
Reader-Interest classification as a method of stock control: the McClellan Legacy	131
<i>Daniel Martínez-Ávila (UNESP) e M. P. Satija (Guru Nanak Dev University)</i>	
Terminology, exhaustivity and specificity: a conceptual relation	140

Isadora Victorino Evangelista, Walter Moreira e João Batista Ernesto de Moraes (UNESP)

Aspects of subject metadata in cataloging photographs.....151
Ana Carolina Simionato (UFSCar)

Documentary languages and audiovisual resources treatment: theoretical and methodological interfaces.....160
Francisnaira Cristina Ravazzi (Diretoria de Ensino) e Walter Moreira (UNESP)

Knowledge Organization and Information Design: a convergence study.....170
Natalia Nakano, Mariana Cantisani Padua e Maria José Vicentini Jorente (UNESP)

Domain Analysis of the first ISKO-Brazil chapter editions.....180
Exequiel Fontans e Mario Barité (Universidad de la Republica)

Teaching in knowledge organization and representation: languages and standards.....196
Cibele Araújo Marques dos Santos, Eduardo de Abreu de Jesus e João Ricardo de Luca (USP)

Genealogy of the concept of Information Science in Brazil: a discursive analysis from founding journals in the area.....206
Larissa de Melo Lima, João Batista Ernesto de Moraes e Daniel Martínez-Ávila (UNESP)

THE APPLY DIMENSION OF KNOWLEDGE ORGANIZATION

Semantic Web Technologies applied to Knowledge Organization: SKOS for the construction and use of decentralized controlled vocabularies.....220
José Eduardo Santarem Segundo (USP) e Caio Saraiva Coneglian (UNESP)

Preliminary analysis on the conversion of classification plans into controlled vocabularies.....229
Luciana Davanzo e Walter Moreira (UNESP)

Thesaurus Reengineering: the Thesagro case.....237
Benildes Maculan (UFMG), Gercina Lima (UFMG), Ivo Pierozzi Jr. (Embrapa) e Leandro Oliveira (Embrapa)

Conceptual Model of Organizational Competitive Intelligence.....246
Thiciane Mary Carvalho Teixeira (UECE) e Marta Lígia Pomim Valentim (Unesp)

Vocabulary control in electronic scientific journals.....256
José Carlos Francisco dos Santos e Brígida Maria Nogueira Cervantes (UEL)

Documentary Languages in archival description.....266
Maria de Fátima Santos de Lima e Francisco Aragão Pedroza da Cunha (UFBA)

Research in knowledge organization systems in DCMI conferences.....277

Felipe Augusto Arakaki, Plácida Leopoldina Ventura Amorim da Costa Santos e Rachel Cristina Vesú Alves (UNESP)

Conceptual models and descriptive representation of information.....288
Elisabete Gonçalves de Souza e Wellington Freire Cunha Costa (UFF)

Nanopublishing Modeling: an experimental approach.....296
Lorena Tavares de Paula e Maria Aparecida Moura (UFMG)

Application of ontologies in the information retrieval process applied in academic settings.....310
Caio Saraiva Coneglian (UNESP), Elvis Fusco (Univem) e José Eduardo Santarém Segundo (USP)

Librarian performance in the subject analysis of theses within the theoretical dimensions of subject cataloging and indexing.....322
Roberta Cristina Dal'Evedove Tartarotti, Paula Regina Dal'Evedove e Mariângela Spotti Lopes Fujita (UNESP)

The contribution of archival identification to knowledge organization in personal archives.....331
Gabrieli Aparecida da Fonseca e Sonia Maria Troitiño Rodriguez (UNESP)

Document analysis and the generative trajectory of meaning in the representation of the archival document.....337
Gilberto Gomes Cândido (UNESP), João Batista Ernesto de Moraes (UNESP) e Deise Sabbag (USP)

Contribution of the Generative Trajectory of Meaning for documentary reading of narrative texts of fiction.....347
Deise Sabbag (USP) e João Batista Ernesto de Moraes (UNESP)

Vocabulary control and its contribution to research: the case of agrarian sciences in Uruguay.....365
Lucía Simón e Mario Barité (Universidad de la República)

Rural cultural goods: an experience report of the taxonomy construction in the context of the historical farms of São Paulo.....373
Mayara Cristina Bernardino (UFSCar), Luciana de Souza Gracioso (UFSCar), Maria da Graça Melo Simões (Universidade de Coimbra) e Luzia Sigoli Fernandes Costa (UFSCar)

Harmonization of CIDOC CRM ontology in the context of archives, libraries and museums.....385
Laís Barbudo Carrasco (UNESP), Silvana Aparecida Borsetti Gregório Vidotti (UNESP) e Phil Manfred Thaller (Universität zu Köln)

Cataloging of interactive resources: analysis of recommendations and practices in international catalogs.....395

Víctor Amor, Daniel Martínez-Ávila (UNESP) e Rosa San Segundo (Universidad Carlos III)

Metadata structure for an image base in pathology.....404

Elisabete Gonçalves Souza, Joice Cleide Cardoso Ennes de Souza e Elan Cardozo Paes de Almeida (UFF)

| 8

Indexing through images: accessibility via OPACs.....416

Rita de Cássia do Vale Caribé e Marcílio de Brito (UnB)

Interactive conceptual maps as a didactic instrument in Information and Knowledge Organization.....435

Marilda Lopes Ginez de Lara, Gabriela Previdello e Nair Yumiko Kobashi (USP)

Domain analysis in Knowledge Organization: exploring thematic and citation relations.....446

Bruno Henrique Alves, Ely Francina Tannuri de Oliveira e Maria Cláudia Cabrini Grácio (UNESP)

A study on domain in knowledge organization through author citation and co-citation analysis.....455

Lidyane Silva Lima, Pollyana Ágata Gomes da Rocha Custódia, Ely Francina Tannuri de Oliveira e Leilah Santiago Bufrém (UNESP)

Classification approaches of the papers submitted to ENANCIB WG2: professional classification and non-professional classification.....463

Walter Moreira e Isabela Santana de Moraes (UNESP)

Information behavior within ISKO - International Society For Knowledge Organization: a study for the 2004-2014 period.....473

Marli Vitor da Silva e Helen de Castro Silva Casarin (UNESP)

Subject representation in Archival Science: an analysis of national and international scientific articles.....481

Graziela Martins de Medeiros, Leolibia Luana Linden, Luciane Paula Vital e Marisa Bräscher (UFSC)

Analysis of methodology in bibliometric studies: a proposal for context indicators.....490

Maria Guiomar da Cunha Frota e Ana Cláudia Ribeiro (UFMG)

Knowledge Organization and Representation: contribution to metric studies501

Cibele Araújo Marques dos Santos (USP)

Relativizing indexes h and g: a study applied to the domain of metric studies508

Deise Deolindo Silva e Maria Cláudia Cabrini Grácio (UNESP)

Domain Analysis in knowledge organization and representation in Archival Science.....	515
<i>Cynthia Maria Kiyonaga Suenaga (UNESP) e Brígida Maria Nogueira Cervantes (UEL)</i>	
The FamilySearch Indexing as a crowdsourcing initiative in the context of knowledge organization.....	526
<i>Paula Carina de Araújo e José Augusto Chaves Guimarães (UNESP)</i>	
A study on actions to make government datasets available in linked open data.....	533
<i>Fernando de Assis Rodrigues e Ricardo César Gonçalves Santana (UNESP)</i>	
Knowledge Organization and industrial heritage in São Paulo: the project Eletromemória.....	544
<i>Vânia Mara Alves Lima (USP), Marcia Cristina e Carvalho Pazin Vitoriano (UNESP) e Cristina Hilsdorf Barbanti (USP)</i>	
THE SOCIAL, CULTURAL AND POLITICAL DIMENSION OF KNOWLEDGE ORGANIZATION	
Reflections on the development of a methodology for subject analysis.....	585
<i>Paula Regina Dal'Evedove (UFPE), Roberta Cristina Dal'Evedove Tartarotti (UNESP) e Mariângela Spotti Lopes Fujita (UNESP)</i>	
Information organization, representation, retrieval and access: (re)configuration of MARC21 Format and BIBFRAME for cultural diversity in digital information environments?.....	592
<i>Zaira Regina Zafalon e Marcela Cristina Nespoli (UFSCar)</i>	
Music, literature and audiovisual: the contributions of knowledge organization (KO) in the intersectional relations between the works by Dorival Caymmi and Jorge Amado.....	598
<i>Fabio Assis Pinho (UFPE), Francisco Arrais Nascimento (UFPE) e Andréa Carla Melo Marinho (UFRGS)</i>	
Mediation in Knowledge Organization Domain.....	605
<i>Mona Cleide Quirino da Silva Farias, Carlos Cândido de Almeida e Daniel Martínez-Ávilla (UNESP)</i>	
The value of information and language in consumer society.....	617
<i>Luciana de Souza Gracioso (UFSCar)</i>	
Academic faculty formation in Archival Description.....	624
<i>Laura Maria do Rego, José Augusto Chaves Guimarães e Natália Bolfarini Tognoli (UNESP)</i>	

PROSPECTS IN KNOWLEDGE ORGANIZATION RESEARCH IN BRAZIL

- Knowledge Organization and the core of Information Science: tales of big data, computing clouds and social networks.....634**
Renato Rocha Souza (FGV), Mauricio Barcellos Almeida (UFMG) e Renata Abrantes Baracho Porto (UFMG)
- Knowledge Organization: research and developments.....644**
Gercina Ângela Borém de Oliveira Lima (UFMG)
- ISKO-Brazil and the research groups in knowledge organization.....663**
Evelyn Goyannes Dill Orrico (UNIRIO)
- Research perspectives on knowledge organization: reflections from PPGCI IBICT-UFRJ academic papers.....671**
Rosalí Fernandez de Souza (IBICT)
- Search perspectives on knowledge organization in Brazil.....683**
Leilah Santiago Bufrem (UNESP/UFPE)

| 10

RESEARCH IN DOCUMENTAL ANALYSIS ON BRAZIL: THE JEAN-CLAUDE GARDIN INFLUENCE

- The search for efficiency in information and knowledge representation – later developments in Gardin’s thought.....700**
Johanna W. Smit (USP)
- Jean Claude Gardin and Document Analysis: trajectory of a semiology of representation.....710**
Maria de Fátima Gonçalves Moreira Tálamo e Giovana Deliberati Maimone (USP)
- Documentary Language from J.-C. Gardin's viewpoint.....721**
Marilda L. G de Lara (USP)
- From Document Analysis to terminology: theoretical and methodological trajectory.....730**
Vania M. A. Lima (USP)
- Ordering Documents: underpinnings and relations with bibliographic classification.....737**
Cristina D. Ortega (UFMG)

INTRODUCTION

Knowledge organization: past, present and future in a cultural and diversified context

| 11



José Augusto Chaves Guimarães
São Paulo State University
guima@marilia.unesp.br

Knowledge Organization - KO, as an area of interdisciplinary knowledge, acts as a convergence space between theoretical and applied approaches and establishes a fruitful dialogue among Logic, Linguistics, Documentation, Communication, Statistics, Psychology, Computer Science, Library Science etc.

Strongly related with Information Science, it has been revealing, over time, a dynamic scientific consolidation that can be explained by its history.

Thus, KO approach presupposes a dynamic conception of knowledge which, once produced, needs to be organized in order to be accessible to society so that this society, once appropriating this knowledge, can generate new knowledge that, in turn, will be socialized and organized in a continuous helical process.

Thus, in the past, i.e., until mid-twentieth century, two concepts stood out in relation to KO configuration: art as a result of a craft and a special talent, result of a technique, and result of the establishment of rules that guide an activity.

In this context, experiences such as the organization of clay envelopes in the Mesopotamian library, Callimachus organization in Alexandria library, Bibliotheca Universalis by Conrad Gessner, Linnaean's taxonomy organization of organisms, Bacon's Harris, Dewey's, Otlet's and Ranganathan's classifications, Poole's index, Kaiser's systematic indexing, PRECIS and POPSI indexing systems and Garfield's citation indexes are evidence of an important development in this field, providing bases for its scientific consolidation.

However, it is especially in the late twentieth century that KO embodied an academic dimension, focused on a more scientific construction. Such fact, had, as landmark, the creation of the International Society for Knowledge Organization - ISKO, as the Society configured, internationally, a space to house investigative advances, scientific dialogue and the dissemination of knowledge produced by an effective system of publications.

Established in July 1989, in Germany, by Ingetraut Dahlberg, from a split occurred within the former Society for Classification, ISKO could constitute, over a quarter century, a complex structure made up of national or regional chapters (Germany/Austria/Switzerland, Brazil, Canada/United States, China, Spain/Portugal, France, India, Iran, Italy, Maghreb, Poland and the United Kingdom), bringing together, currently, more than 600 researchers, in order to "promote research, development and applications of knowledge organization systems that advance the philosophical, psychological and semantic approaches for ordering knowledge, provide the means of communication and networking on knowledge organization for its members, function as a connecting link between all institutions and national societies, working with problems related to the conceptual organization and processing of knowledge"(ISKO).

In this context, ISKO holds a set of publications of international scope in order to disseminate scientific knowledge generated there: the bimonthly journal Knowledge Organization; the series Advances in Knowledge Organization, with the proceedings of ISKO international conferences held in even years in different countries worldwide; and the proceedings of ISKO chapters' conferences held in odd years.

Concerning its theoretical construction, some aspects deserve special attention in the contemporary context of KO, such as the issue of knowledge, its records and its socialization (BUCKLAND 1991, 1997; FROHMANN, 2004, 2009), the epistemological configuration of KO as an area (DAHLBERG, 1993, 1995; HJØRLAND 1993, 2003, 2008, TENNIS, 2008; LÓPEZ HUERTAS, 2008; GNOLI, 2008, 2011; DOUSA, 2009; SMIRAGLIA, 2009, 2012; ZINS, 2004; THELLEFSEN & THELLEFSEN 2004), cultural aspects that have impacted the area (HUDON, 1997, 1999; OLSON, 2002; BEGHTOL, 2002, 2005; GARCÍA GUTIÉRREZ, 2002; HJØRLAND, 2008; MILANI & GUIMARAES, 2011; CAMPBELL, 2009), and applied issues, involving aspects related to representation languages and technological impact (ANDERSEN, 2002; BEAK & OLSON, 2011; BUCHEL & HILL, 2009; EDWARDS, 2011; GREEN & FALLGREN, 2007; GREEN, 2011; HOFFMAN, 2009; KIPP, 2007).

In this scenario, ISKO-Brazil, since its first congress in 2011, in Brasilia, then the second congress, in Rio de Janeiro in 2013 and, in this opportunity, in its third congress in Marilia, in 2015, has been addressing discussions from a thematic triad composed by: an epistemological dimension, focused on the consolidation of theories, methodologies, paradigms, schools of thought; an applied dimension, strongly related to technological advances and development new of instruments; and a cultural dimension, focused on the contexts, communities and individuals who interact in KO.

In this three-dimensional perspective, some challenges can be envisioned, such as the need, in the epistemological dimension, for evidencing different theoretical and methodological currents that permeate the area, as well as their intersections, (schools of thought in KO); in the applied dimension, avoiding information waste and developing friendlier tools; and, in the cultural dimension, avoiding prejudice, proselytism and cultural domination as well as promoting compatibility between the need for global communication and respect to local issues.

Furthermore, interesting perspectives are presented, such as the consolidation of KO as knowledge domain and the configuration of investigative inter and transdisciplinary spaces in the area (epistemological dimension), the promotion of interoperability and more efficient, effective and culturally significant information retrieval (applied dimension), and in the cultural dimension, the establishment of a transcultural ethics of mediation (GARCÍA GUTIERREZ, 2002), permeated by cultural warrant (Beghtol, 2002, 2005) and respect for knowledge domains (THELLEFSEN & THELLEFSEN, 2004).

Such aspects lead us to believe in a promising future for the area of KO, with the possibility of creating new chapters of ISKO (national and regional), meaning new and diversified areas of scientific dialogue, intensification and multiplication of inter-institutional and international scientific partnerships, the consolidation of this knowledge domain as nuclear investigative space in Information Science, and the growth (in scope and in depth) of research in the area, permeated by the need for constant and sharp criticism inherent to dynamic spaces.

In this opportunity, ISKO-Brazil promotes its third congress, under the theme *Knowledge Organization and Cultural Diversity*, an aspect that finds ballast in the reality experienced by our country and the concerns arising from it, and also sets the ground for the 14th ISKO Conference, to be held in Rio de Janeiro, in September 2016,

under the theme *Knowledge Organization for a sustainable world: challenges and perspectives for cultural, scientific and technological sharing in a connected society*.

This ISKO-Brazil Congress counts with a set of seventy-five papers, including conferences and communications addressing the three dimensions previously mentioned, in addition to bring a reflection on the research perspectives in KO in Brazil, and on the influence of Jean-Claude Gardin research in document analysis in Brazil.

| 14

In this context, we have to register the role of the Information Science Graduate Program of UNESP (São Paulo State University), as co-promoter institution, and also the important support received from FAPESP (São Paulo Research Foundation), CNPq (Brazilian national Council on Research and Scientific and Technological Development), Capes (Brazilian Higher Education Development Foundation), the FUNDUNESP (UNESP Development Foundation) and VUNESP, organizations that made the event possible.

We also highlight the publication of four abstracts, selected as the best congress papers, which will be published in full in the journal *Knowledge Organization*, as well as three other papers that received ISKO-Brazil award, to be published in full in the *Brazilian Journal of Information Science - BRAJIS*.

We expect, with this work, to contribute to scientific sedimentation of KO in Brazil as well as the vertical integration of its approaches, providing greater visibility and scientific dialogue.

References

The references was made following the ABNT rules.

ANDERSEN, J. Communication technologies and the concept of knowledge organization. *Knowledge Organization*, Wuzburg, v. 29, p. 29-39, 2002.

BEAK, J.; OLSON, H. A. Analysis of metadata schemas for children's libraries. In: *North American Symposium on Knowledge Organization*, 3, 2011, Toronto, Proceedings from North American Symposium on Knowledge Organization, p. 13-22. Available: <http://iskocus.org/nasko2011-proceedings.php>.

BEGHTOL, C. A proposed ethical warrant for global knowledge representation and organization systems. *Journal of Documentation*, London, v. 58, n. 5, p. 507-532, 2002.

BEGHTOL, C. Ethical decision-making for knowledge representation and organization systems for global use. *Journal of the American Society for Information Science and Technology*, New York, v. 56, n. 9, p. 903-912, 2005.

BUCHEL, O.; HILL, L. Treatment of Georeferencing. in Knowledge Organization Systems: North American Contributions to Integrated Georeferencing. In: *North American Symposium on Knowledge Organization*, 2, 2009, Syracuse, NY, Proceedings from North American Symposium on Knowledge Organization, p. 47-57. Available: <http://iskocus.org/nasko2009-proceedings.php>.

BUCKLAND, M. What is a document? *Journal of the American Society for Information Science*, Washington, v.48, n.9, p.804-809, 1997.

BUCKLAND, M. Information as thing. *Journal of the American Society of Information Science*, Washington, v. 42, n. 5, p. 351-360, 1991.

CAMPBELL, D. G. Tensions between language and discourse in north american knowledge organization. In: *North American Symposium on Knowledge Organization*, 2, 2009, Syracuse, NY, Proceedings from North American Symposium on Knowledge Organization, p. 10-16. Available: <http://iskocus.org/nasko2009-proceedings.php>.

DAHLBERG, I. Knowledge organization: its scopes and possibilities. *Knowledge organization*, Frankfurt, n. 20, p. 211-222, 1993

DAHLBERG, I. Current trends in Knowledge Organization. In: *Organización del conocimiento en sistemas de información y documentación*, 1, 1993, Madrid, Actas del I Encuentro de ISKO-España, Zaragoza, Librería General, 1995, p. 7-25.

DOUSA, T.M. Classical Pragmatism and its Varieties: On a Pluriform Metatheoretical Perspective for Knowledge Organization. In: *North American Symposium on Knowledge Organization*, 2, 2009, Syracuse, NY, Proceedings from North American Symposium on Knowledge Organization, p.:1-9. Available: <http://iskocus.org/nasko2009-proceedings.php>.

EDWARDS, C. (2011). Soviet classifications and situated knowledges. In North American Symposium on Knowledge Organization, 3, 2011, Toronto, Proceedings from North American Symposium on Knowledge Organization, p. 13-22. Available: <http://iskocus.org/nasko2011-proceedings.php>.

FORHMANN, B. *Deflating information: from science studies to documentation*. Toronto: University of Toronto Press, 2004.

FROHMANN, B. Revisiting "What is a document". *Journal of Documentation*, London, v.6, n.2, p.291-303, 2009.

GARCÍA GUTIÉRREZ, A. Knowledge organization from a "culture of the border": towards a transcultural ethics of mediation. In: LÓPEZ-HUERTAS, M. J. (Ed.). Challenges in knowledge representation and organization for the 21st century: integration of knowledge across boundaries. Würzburg: ERGON-Verlag, 2002. p. 516-522.

GNOLI, C. Ten long-term research questions in knowledge organization. *Knowledge organization*, Würzburg, v.35, n.2/3, p.137-149, 2008.

GNOLI, C. Ontological foundations in KO: the theory of integrative levels applied in citation order. *Scire*, Zaragoza, v. 17, n. 1, p. 29-34, 2011.

GREEN, R.; FALLGREN, N. Anticipating new media: A faceted classification of material types. In *North American Symposium on Knowledge Organization*, 1, Toronto, 2007, *Proceedings North American Symposium on Knowledge Organization*, p. 87-99. Available: <http://iskocus.org/nasko2007-proceedings.php>.

GREEN, R. See-also relationships in the Dewey Decimal Classification. In *North American Symposium on Knowledge Organization*, 3, 2011, Toronto, *Proceedings from North American Symposium on Knowledge Organization*, p. 41-51. Available: <http://iskocus.org/nasko2011-proceedings.php>.

HJØRLAND, B. Deliberate bias in knowledge organization? In: *Culture and identity in knowledge organization*, 11, Montréal, 2008, *Advances in knowledge organization*, Würzburg: Ergon Verlag, 2008. p. 256-261.

HJORLAND, B. Fundamentals of knowledge organization. *Knowledge Organization*, Würzburg, v. 30, n. 2, p. 87-111, 2003.

HJORLAND, B. Nine principles of knowledge organization. In: *Knowledge organization and quality management*, 4, Copenhagen, *Advances in Knowledge Organization*, Frankfurt/main: Indeks, 1994. p. 91-100.

HJORLAND, B. What is knowledge organization (KO). *Knowledge organization*, Frankfurt, v. 35, n. 2/3, p. 86-101, 2008.

HOFFMAN, G.L. Applying the User-Centered Paradigm to Cataloging Standards in Theory and Practice: Problems and Prospects. In: *North American Symposium on Knowledge Organization*, 2, 2009, Syracuse, NY, *Proceedings from North American Symposium on Knowledge Organization*, p. 27-34. Available: <http://iskocus.org/nasko2009-proceedings.php>.

HUDON, M. Accessing documents and information in a world without frontiers. *The Indexer*, London, v. 21, n. 4, p. 156-159, 1999.

HUDON, M. Multilingual thesaurus construction: integrating the views of different cultures in one gateway to knowledge and concepts. *Knowledge Organization*, Würzburg, v. 24, n. 2, p. 84-91, 1997.

ISKO's mission. Available at: <http://www.isko.org/about.html>. Access at: 02.09.2015.

KIPP, M. E. I. Tagging for health information organisation and retrieval. In *North American Symposium on Knowledge Organization*, 1, Toronto, 2007, *Proceedings North American Symposium on Knowledge Organization*, p. 63-74. Available: <http://iskocus.org/nasko2007-proceedings.php>.

LÓPEZ-HUERTAS, M. J. Some current research questions in the field of knowledge organization. *Knowledge Organization*, Würzburg, v. 35, n. 2/3, p. 113-136, 2008.

MILANI, S. O.; GUIMARÃES, J. A.C. (2011). Biases in knowledge representation: an analysis of the feminine domain in Brazilian indexing languages. In North American Symposium on Knowledge Organization, 3, 2011, Toronto, Proceedings from North American Symposium on Knowledge Organization, p. 94-104. Available: <http://iskocus.org/nasko2011-proceedings.php>.

OLSON, H. A. *The power to name: locating the limits of subject representation in libraries*. Dordrecht: Kluwer Academic Publishers, 2002.

SMIRAGLIA, R. P. Modulation and Specialization in North American Knowledge Organization: Visualizing Pioneers In: North American Symposium on Knowledge Organization, 2, 2009, Syracuse, NY, Proceedings from North American Symposium on Knowledge Organization, p. 35-46. Available: <http://iskocus.org/nasko2009-proceedings.php>.

SMIRAGLIA, R.P. Epistemology of domain analysis. In: SMIRAGLIA, R. P.; LEE, H-L. (eds), *Cultural frames of knowledge*. Würzburg: Ergon-Verlag, 2012, p. 111-24.

TENNIS, J. T. Epistemology, theory and methodology in Knowledge Organization: Toward a Classification, Metatheory, and Research Framework. *Knowledge organization*, Frankfurt, v. 35, n. 2/3, p. 102-112, 2008.

THELLEFSEN, T. L; THELLEFSEN, M. M. Pragmatic semiotics and knowledge organization. *Knowledge Organization*, Würzburg, v. 31, n. 3, p. 177-187, 2004.

ZINS, C. Knowledge organization: an epistemological perspective. *Knowledge organization*, Würzburg, v. 31, p. 49-54, 2004.

- Organizing Committee: Vera Dodebei
- Secretary: Renato Rocha de Souza, Natália Bolfarini Tognoli, Daniel Martinez-Ávila
- Treasurers: Evelyn Goyannes Dill Orrico, Silvana B. G. Vidotti

The Epistemological Dimension of Knowledge Organization

Epistemic, Ontic, Axiologic, and Praxic Constructs in Knowledge Organization Research



Joseph T. Tennis
University of Washington
jtennis@uw.edu

Introduction

The work of knowledge organization requires a particular set of tools. For instance we need standards of content description like Anglo-American Cataloging Rules Edition 2, Resource Description and Access (RDA), Cataloging Cultural Objects, and Describing Archives: A Content Standard. When we intellectualize the process of knowledge organization – that is when we do basic theoretical research in knowledge organization we need another set of tools. For this latter exercise we need *constructs*. Constructs are ideas with many conceptual elements, largely considered subjective. They allow us to be inventive as well as allow us to see a particular point of view in knowledge organization. For example, Patrick Wilson’s ideas of exploitative control and descriptive control, or S. R. Ranganathan’s fundamental categories are constructs. They allow us to identify functional requirements or operationalizations of functional requirements, or at least come close to them for our systems and schemes. They also allow us to carry out meaningful evaluation.

What is even more interesting, from a research point of view, is that constructs once offered to the community can be contested and reinterpreted and this has an affect on how we view knowledge organization systems and processes. Fundamental categories are again a good example in that some members of the Classification Research Group (CRG) argued against Ranganathan’s point of view. The CRG posited more fundamental categories than Ranganathan’s five, Personality, Matter, Energy, Space, and Time (Ranganathan, 1967). The CRG needed significantly more

fundamental categories for their work.¹ And these are just two voices in this space we can also consider the fundamental categories of Johannes Kaiser (1911), Shera and Egan, Barbara Kyle (Vickery, 1960), and Eric de Grolier (1962). We can also reference contemporary work that continues comparison and analysis of fundamental categories (e.g., Dousa, 2011).

In all these cases we are discussing a construct. The fundamental category is not discovered; it is constructed by a classificationist. This is done because it is useful in engaging in the act of classification. And while we are accustomed to using constructs or debating their merit in one knowledge organization activity or another, we have not analyzed their structure, nor have we created a typology. In an effort to probe the epistemological dimension of knowledge organization, we think it would be a fruitful exercise to do this. This is because we might benefit from clarity around not only our terminology, but the manner in which we talk about our terminology. We are all creative workers examining what is available to us, but doing so through particular lenses (constructs) identifying particular constructs. And by knowing these and being able to refer to these we would consider a core competency for knowledge organization researchers.

In the next section we lay out our preliminary typology of constructs and in so doing identify some attributes of the constructs under investigation. Of course we cannot, in the space of this paper, describe all constructs used in knowledge organization. However, this is a start to a conversation that we hope improves our understanding of the philosophical commitments we make in the practice of knowledge organization research.

Constructs and their Kinds

It is possible for us separate out constructs along a variety of axes. We can consider their philosophical nature, and ask whether they about things in knowledge organization or about how we know about some aspect of knowledge organization. The former is an *ontic construct* the latter is an *epistemic construct*. A third category

¹ For example, in the classification of science Brian Vickery proposes: Substance (product), organ, constituent, structure, shape, property, object of action (patient, raw material), action, operation, process, agent, space, and time. See Vickery (1960) p. 23. And see also Austin (1968), Foskett (1974), Mills and Broughton (1977).

of philosophical construct can be identified as well. *Axiologic constructs* are those that focus on worth – both aesthetic and ethical aspects of knowledge organization. One final construct we consider here is the *praxic construct*. These are constructs that deal with particular actions and activities in knowledge organization. In the following section we provide examples and describe further the attributes of these constructs.

Ontic Constructs: What there is in the world of knowledge organization

As mentioned above, fundamental categories are constructs we use in order to create analytico-synthetic classification (*vide* Ranganathan, 1967). Fundamental categories are *things* we construct in order to do the work at hand. What is interesting about ontic constructs is how researchers agree or disagree with their definitions and boundaries. One such ontic construct is the *subject*, which we have represented as a real-world object that can be faithfully represented in our systems. However the nature of subjects is, in the contemporary literature, often discussed from a pragmatic stance – that of an informed judgment, which has a sometimes-tense relationship to realist ontologies. For example, there is a fundamental disagreement in the literature about the nature of subjects (Kučerová, 2014). By acknowledging these characteristics, we can begin to refine our assumptions and our knowledge of subjects as mobilized in our information systems.

Ontic constructs form the largest set of constructs in knowledge organization literature. We have for the past one-hundred and fifty years or so, made *things* that exist in the universe of knowledge organization. The other constructs, while not immediately recognizable as things are just as valuable, because they make up a different kind of construct. The work of the remaining three constructs is not about what *is* in the universe of knowledge organization, but rather what in the universe of knowledge *has worth, allows us to know, and what we do*. We will first talk about what in the universe of knowledge organization has worth.

Axiologic Constructs

With the rising concern about the ethics of knowledge organization systems, we have seen a rise of axiologic constructs. These constructs about the value of knowledge organization systems, both ethical and aesthetic, interrogate the worth of our work in knowledge organization. If we are adding value, how do we assess that value? Classification as rhetorical device, a position advanced by Melanie Feinberg

(Feinberg 2011), is one way to look at knowledge organization systems in an aesthetic way. The value presented here is in the classification scheme's ability to persuade me, the user (or perhaps reader) of the scheme of the author's position.

There are also ethical constructs that have surfaced in our contemplation about the potential harm our systems might cause to our users. This line of research is indebted to Sanford Berman (1971) and continues today with contemporary scholarship, with Melissa Adler (2013), Fabio Pino (2010), Melody Fox (2012) following on substantial contributions by Hope Olson (2002) Jonathan Furner (2008), and Clare Beghtol (2002).

Constructs in this space are a bit different than the ontic constructs above. In many cases we have to do some creative naming in order to isolate them from the texts. For example, Guimarães et al., identify ethical values and ethical problems in knowledge organization, while acknowledging a wider concern in ethical issues in contemporary global society (Guimarães et al., 2008). In order to make ethical values and ethical problems into constructs, we might need to compare these lists with others and squarely define what we mean by values and problems. Would Guimarães et al., agree with other thinkers in the field as to whether something is an ethical value or problem for knowledge organization? We would only know this if we had a definition and extension of the construct that can persist through time and across contexts. If there is disagreement among researchers in knowledge organization, then we again see the fruitfulness of constructs – by being explicit about what we are constructing we can come to clarity about the substance of our discussions and, by extension, the way we see knowledge organization systems can be improved.

The bottom line with regard to *axiologic constructs* is that we must be able to articulate value and its corollaries, ethics and aesthetics where appropriate in knowledge organization systems. We can do this by understanding of the ideas present in the practice and reflection of knowledge organization as ideas with conceptual elements – i.e., constructs.

Epistemic Constructs

Ontic and axiologic constructs say what is and what has value in knowledge organization. However, a core concern in knowledge organization is epistemology and epistemological issues. These issues at their most basic ask *how do we know* some aspect of knowledge or knowledge organization. In an applied way, knowledge

organization has also crafted epistemology and epistemic concerns, to be the purview of theories of knowledge. That is, it is not only concerned with how we know, but clarifying our assumptions versus our intentions with how we know. Further, investigating and manufacturing epistemic constructs can give rise to other constructs, particularly ontic constructs

David Blair is one fundamental thinker in this space when he argues for a particular epistemic stance on language and information retrieval (Blair, 1990; 2005). In the case of Blair, we have the epistemic construct of *language as speech act* where we claim to know the ontic construct (language) in a particular way (as a speech act). This is a refinement of the assumption that we are dealing with the pragmatic view of the universe of language. The relationship between the universe of language and the universe of knowledge is a core concern for knowledge organization, and by making explicit what that relationship is, if they are indeed separate and separable constructs is the business of epistemic constructs we deploy in knowledge organization research.

For example in reference theory, we construct an epistemic construct of the concept triangle with its three parts: symbol, thought, referent (Ogden and Richards, 1923). This is an epistemic construct because it points to how we should know the relationship between language and knowledge.

Named different schools of thought can be seen as epistemic constructs too. So rationalist, empiricist (Hjørland, 2005), analogic (Lee, 2012), indigenous imagining (Duarte and Belarde-Lewis, 2015), are all constructs that allow us views into what ontic constructs are available to us, and informs our axiologic constructs. For example, the concept of harmony might not be important to rationalist or empiricist constructs, but that cannot be said of analogic or indigenous imagining – where what is known is couched in a deep relationship to an ethos of harmonious being in relation to the cosmos. Again we can mention Thomas Dousa because work fits here as well. He is very concerned with rendering clear the meaning behind labels. For example he has probed what it means to say one is taking a pragmatist's stance in relation to knowledge organization (Dousa, 2010).

The final example presented here is the concept of naïve classification and professional classification (Beghtol, 2003). What we know as classification and how we conceptualize its place in the universe of knowledge organization research is an epistemic construct. This is born out by the debate over what was actually under discussion in Beghtol's article (Nicholaisen and Hjørland, 2004; Jacob, 2010). We

have a tradition of naming these after the authors, and that seems reasonable to continue if not further encourage.

Praxic Constructs

The final constructs are those that deal with the act and action of knowledge organization. Indexing, classification, domain analysis, cataloguing are all actions and each of them can be defined and described in a variety of ways. This makes them constructs. Further, any subdivision of these acts can be considered constructs, named and therefore recognized as objects of investigation. Jens-Erik Mai has described a three step indexing process (Mai, 2003). He subdivided subject indexing into (1) document analysis, (2) subject description process, (3) subject analysis process (Mai, 2003 p. 595). Each of these is an action and he has offered them as praxic constructs whereby we can probe further (in agreement or disagreement with others) into how indexing occurs.

Interestingly we have thought tagging on social media websites to be similar to indexing (Tennis, 2006; Kip 2014). The assumption here seems to be a bit at the intersection between epistemic and praxic constructs because we are assuming one way of identifying indexing

Next Steps

If constructs are useful for knowledge organization research to consider we have a number of next steps we can take. The first is to identify the population of constructs that exist in the knowledge organization literature. The next is to figure out which are the same constructs under different names and which are different constructs under the same name. And finally we need to consider which are core to our present concerns in knowledge organization, which are orthogonal or corollary – which are not core, but play important roles at times in our discussions of knowledge organization research, and which are outmoded, perhaps too rooted to extinct technology or outdated functional requirements. An example of core constructs might be facets, ethical values, pragmatism, and indexing all mentioned above. An orthogonal construct might be indigenous imagining where we are concerned with involving Native American people in designing knowledge organization systems or a separate construct for the action of tagging in the context of Twitter. Outmoded constructs could always come back based on their value or utility, but we do not as a

matter of routine talk about *abstract classification* the way S. R. Ranganathan did in the 1960s (Ranganathan, 1967). It is possible that we might talk about abstract classification sometime in the future, but we are no longer actively working on this in this community.

Conclusion

This paper has offered the idea of a family of constructs in knowledge organization research. I have done this in an effort to probe the epistemological dimension to knowledge organization. What we know and how we talk about what we know in knowledge organization is deeply rooted in two camps: philosophy and practice. We have, through the natural course of our work, naively created constructs to help us work with our universe of knowledge organization. Classes, facets, ethical problems, empiricism, the act of cataloguing are all constructs borrowed from philosophy or practice and bound together in a particular way we use language in knowledge organization. They are ours as much as they are anyone else's. However, we must appreciate our particular needs in naming these in our literature. We would argue that we need these to be constructs. We need to deploy these to create, maintain, and evaluate knowledge organization systems. We have a particular use for these ideas that we have seen in the literature of philosophy and the everyday life of practice.

Knowledge organization systems are the results of a deeply philosophical and imminently practical activity of making the world's recorded knowledge available to users. We have approached this problem by creating our systems out of our understanding of knowledge and users, which includes language, technology, context, and values. By probing deeper into constructs as part of the epistemological dimension we move toward a systematic understanding of how we have crafted the relationship between philosophy and practice and how we have created an inventory of our contributions to helping users access the universe of knowledge.

References

Adler, M. A. (2013). The ALA Task Force on Gay Liberation: Effecting Change in Naming and Classification of LGBTQ Subjects. *Advances In Classification Research Online*, 23(1), pp. 1-4

Austin, D. (1968). *Fields, categories and general systems theory: Report to the Classification Research Group*. (London: BNB).

- Berman, S. (1971). *Prejudices and Antipathies* (Metuchen, NJ: Scarecrow Press).
- Beghtol, C. (2002). "A proposed ethical warrant for global knowledge representation and organization systems." In *Journal of Documentation* 58(5): 507-532.
- Beghtol, C. (2003). "Classification for Information Retrieval and Classification for Knowledge Discovery: Relationships between "Professional" and "Naïve" Classifications." In *Knowledge Organization* 30(2): 64-73
- Blair, D. C. (1990). *Language and Representation in Information Retrieval*. (Elsevier).
- Blair, D. C. (2005). Information retrieval and the philosophy of language. In *Annual Review of Information Science and Technology*.
- Dousa, T. M. (2010). "Classical Pragmatism and its Varieties: On a Pluriform Metatheoretical Perspective for Knowledge Organization." *Advances in Knowledge Organization Research*
- Dousa, T. M. 2011. Concretes, countries, and processes in Julius O. Kaiser's Theory of Systematic Indexing: A case study in the definition of general categories. In Smiraglia, Richard P., ed. *Proceedings from North American Symposium on Knowledge Organization*, Vol. 3. Toronto, Canada, pp. 160-173.
- Duarte, M. E. and Belarde-Lewis, M. (2015). "Imagining: Creating Spaces for Indigenous Ontologies." In *Cataloging and Classification Quarterly* 53(5-6): 677-702.
- Feinberg, M. (2011). Feinberg. Compiler to author: a process for designing rhetorically aware document collections. *Journal of the American Society for Information Science and Technology* 62(9): 1784-1796.
- Foskett, D. J. (1974). *Classification and indexing in the social sciences*. (London: Butterworths).
- Fox, M.J. & Olson, H.A. (2012). Feminist epistemologies and knowledge organization. In Hur-li Lee and Richard Smiraglia (eds.), *Cultural frames of knowledge*. Würzburg: Ergon Verlag. 79-98.
- Furner, J. (2008). Dewey deracialized: A critical race-theoretic perspective. In *Knowledge Organization* 43(3): 144-168.
- Grolier, E. de. (1962). *A study of general categories applicable to classification and coding in documentation*. (Paris: UNESCO).
- Guimarães, J. A. C., Fernández-Molina, J. C., Pinho, F. A., Milani, S. O. (2008). Ethics in the Knowledge Organization Environment: An Overview of Values and Problems in the LIS Literature." In *Advances in Knowledge Organization* :361-366.
- Hjørland, B. (2005). "Empiricism, rationalism and positivism in library and information science." In *Journal of Documentation* 61(1): 130-155.
- Jacob, E. K. (2010). "Proposal for a Classification of Classifications built on Beghtol's Distinction between "Naïve Classification" and "Professional Classification"" In *Knowledge Organization* 37(2): 111-120.
- Kaiser, J. (1911). *Systematic Indexing*. (Isaac Pitman & Sons).

Kipp, M. (2014). "Understanding the retrieval effectiveness of collaborative tags and author keywords in different retrieval environments: An experimental study on medical collections." In *Journal of the Association for Information Science and Technology* 65(3): 483-500.

Kučerová, H. (2014). "What Do we Analyze during Subject Analysis of Documents? The Concept of Aboutness in Knowledge Organization." In *Revue of Librarianship* 25(1): 38-57.

Lee, H-L. (2012). "Epistemic Foundation of Bibliographic Classification in Early China: A Ru Classicist Perspective." In *Journal of Documentation* 68(3): 378-401.

Mai, J-E. (2003). "Semiotics and indexing: An analysis of the subject indexing process. In *Journal of Documentation* 57(5): 591-622.

Mills, John, and Vanda Broughton. (1977). *Bliss bibliographic classification. Second edition.* (London: Butterworths).

Nicholiasen, J. and Hjørland, B. (2004). "A rejoinder to Beghtol." In *Knowledge Organization* 31: 199-201.

Ogden, C. K. and Richards, I. A. (1923). *The Meaning of Meaning: A Study of the Influence of Language upon Thought and of the Science of Symbolism.* (Trench, Trubner & Co)

Olson, H. A. (2002). *Power to name.* (Springer).

Pinho, F. A. (2010). Aspectos éticos em representação do conhecimento em temáticas relativas à homossexualidade masculina: uma análise da precisão em linguagens de indexação brasileiras. Dissertation for UNESP.

Ranganathan, S. R. (1967). *Prolegomena to Library Classification.* 3rd Edition. (Bombay: Asia Publishing House).

Tennis, J. T. (2006). "Social Tagging and the Next Steps for Indexing: Fordist Reflexivity and Intertextuality." In *Advances in Classification Research.* vol. 17. (Austin, Texas). Available: <http://dlist.sir.arizona.edu/1726/>

Vickery, B. C. (1960). *Faceted classification: A guide to construction and use of special schemes.*

Approaches and paradigms in knowledge organization



Leila Cristina Weiss
Federal University of Santa Catarina
leilacw@gmail.com



Marisa Bräscher
Federal University of Santa Catarina
marisa.brascher@gmail.com

1 Introduction

There are different proposals of classification by theoretical and methodological currents in Knowledge Organization (KO) which, in general, are the result of epistemological discussions on the field, which is interdisciplinary and receives contributions from areas such as Information Science, Philosophy, Linguistics, Semantics, Computer Science, among others.

One of the classification proposals is the one by Hjørland (2007), and developed from the analysis of paradigms in semantics, due to the strong influence of this area on the KO. From this perspective, we distinguish the pragmatic and positivist approaches in KO.

Pragmatic, positivist, operationalist, referential, instrumental, empiricist, rationalist and realistic are some names for epistemological positions, according to Tennis (2008). The proposed classification of theoretical and methodological currents present in KO and developed by Gnoli (2008), however, does not refer to any of these epistemological positions, but to the ontological approach and epistemological approach, which the author considers the two basic approaches in KO.

Due to the importance of proposals by Hjørland (2007) and Gnoli (2008), reflected in recent research and discussion in KO, this paper comparatively analyzes these two theoretical and methodological currents in KO. Hjørland (2007) distinguishes KO paradigms into pragmatic and positivist, and Gnoli (2008), distinguishes the KO approaches into ontological and epistemological. We describe each of the proposals and point out similarities, differences and complementarities between them.

2 Pragmatic and positivist paradigm in KO

We begin the discussion on Hjørland's proposal (Hjørland, 2007) for the classification of theoretical and methodological currents in research on KO, from a previous work by the same author, published in 2003, which describes the presence of the empiricist, rationalist, historicist and pragmatic epistemological positions in KO.

Hjørland (2003) points out that KO methods are related to epistemological assumptions, among which he highlights empiricism, rationalism, historicism and pragmatism. According to the author, in KO some articles and systems rely mostly on empirical generalizations, such as for example, systems based on the word frequency measurement. Other articles and systems mostly adopt rational rules and deductions (although often ignoring empirical questions). In this case, the author considers that the analytical-faceted systems on Ranganathan's tradition provide the best examples. A third type of system is based on the study of the evolution of knowledge areas (historicism). To some extent, the systems with disciplinary bases, such as the Dewey Decimal Classification (DDC), would be in accordance with that ideal.

On the other hand, the updating of this system is more influenced by the principles of analytical-faceted tradition (as discussed by Miksa, 1998), and therefore, in a sense, DDC also has an empirical basis. The studies and systems based on pragmatism are developed based on the analysis of objectives, values and consequences. The author points out that this paradigm in KO lies within the realistic position, as the social constructions produced only from personal interests or interests from researchers in the field, when confronted with reality, may prove inconsistent. Thus, the pragmatic method does not oppose to aspects of empiricism, rationalism and historicism. However, isolated evidence is not sufficient to support KO processes and the necessary evidence is related to the objectives and to human activities. (Hjørland, 2003).

In this direction, Tennis (2008) argues that epistemology involves a decision on which knowledge is 'valid' for KO, and what can be considered as a source of acceptable evidence of this knowledge. Therefore, "Epistemology in KO results in epistemic stance that outlines knowledge claims" (TENNIS, 2008, p.105)

These epistemological assumptions about language are discussed by Hjørland (2007), who considers semantics as an area of extreme importance to KO. According to the author, different approaches to KO imply different views on semantics.

Peregrin (2004) highlights two dominant paradigms in semantics: the pragmatic and the positivist. In the positivist paradigm, language is portrayed as a means of representing the world and reality. In the pragmatic paradigm, language is seen as a means of interaction and the meaning of an expression is not the entity represented by it, but its function in the interaction. Based on Peregrin (2004), Hjørland (2007) distinguishes paradigms in KO into pragmatic and positivist, unlike how he had defined the paradigms in 2003.

Francelin and Kobashi (2011), similarly, distinguish the approaches to the concept in KO into positivist and pragmatic. In their study, based on the Brazilian literature published in scientific articles, these researchers conclude that "the theoretical force lines are underpinned, fundamentally, in Aristotelian logic, best represented by Categories, and the principle of language fixing or concepts through judgments and true propositions, basic characteristics of analytical philosophy and logical positivism "(Francelin; Kobashi, 2011, p.221).

On the other hand, according the authors mentioned above, it is possible to find pragmatic underpinnings on the use of language and concepts, especially in articles that "analyze the concept from the perspective of philosophy of language, semiotics, linguistics, communicative theory of terminology and socioterminology ". (Francelin; Kobashi, 2011, p.222).

Both Hjørland (2007) and Francelin and Kobashi (2011) point out that, in general, research on KO does not explicitly indicate in which theory or philosophical current they are based. But indicate a strong inclination to positivism, and Francelin and Kobashi (2011) also state that they often use foundations of this theoretical current, such as formal logic and Aristotelian categories and logic, uncritically. Hjørland (2009) addresses theory of concept under the epistemological perspective and considers the pragmatic and historicist views as the most productive for the development of KO.

Thus, for the analysis of the purposes of this study, the considered proposed classification was Hjørland's (Hjørland , 2007), who distinguishes the paradigms in KO only into pragmatic and positivist. In general, an adaptation of his original proposal published in 2003 can be considered, since in its proposal for 2007, the rationalist and

empiricist paradigms would fit in the positivist paradigm, and the historicist and pragmatic paradigms, would fit into pragmatic paradigm.

3 Ontological and epistemological approaches to KO

Gnoli (2008) begins his description of ontological and epistemological approaches to KO from considerations on Ontology and Epistemology, both subfields of philosophy. In this context, he states that Ontology

[...] concerns the nature of the known things, especially in terms of the general categories to which they may belong. Issues like the subdivision of a class into kinds and parts or the acknowledgement that a given concept consists of a process or a static entity, are ontological. Epistemology, instead, is about how humans know the world through their sense organs, and how they process knowledge according to categories both innate and culturally biased (GNOLI, 2008, P. 139).

In the author's view, "knowledge is both epistemological and ontological, as it passes through human perception by its very nature but also refers to real objects of the world having some intrinsic structure." (Gnoli, 2008, p.139). Documented knowledge, information, carries these characteristics. Redón Rojas (1996) considers that information is constructed, and not simply invented (subjective) or found (objective). Information is not a simple invention, its elaboration is supported on certain raw material (objective elements which are found on things); nor is it something ready which and only needs to be discovered, as it is necessary to man's action. Men transform and are transformed by the informational world in a constantly evolving process. In this concept of information, the dialectic of the subject with the world around him is taken into consideration.

Also in Gnoli's view (Gnoli, 2008), in the ontological approach, the focus lies on the world and the supposed objective reality, while in the epistemological approach, the focus lies on the subject and the social context, as in the example:

Dewey's main classes follow an epistemological sequence, going back to Francis Bacon, as they are listed according to basic forms of the human intellect producing them, like reason, imagination, and memory; UDC main classes are also epistemological, as they are derived from Dewey. Other systems, like the Bliss Classification, the Broad System of Ordering, and the Information Coding Classification, base the sequence of their main classes on a supposedly natural sequence of

increasing specificity and complexity of the known objects, hence they are primarily ontological (GNOLI, 2008, P. 139)

Following with the distinction between these two approaches in KO, Gnoli (2008), cites Dahlberg (1974; 1978) and Robert Poli, who reference the philosopher Nicolai Hartmann, as representative of the ontological approach, and Hjørland and Albrechtsen (1995) as representatives of the epistemological approach. Researchers in KO who develop their research from the perspective of the epistemological approach, according Gnoli (2008, p.139), "often take examples from language and its cultural relativity, and quote philosophers like Wittgenstein and the American pragmatists".

| 32

The author does not advocate either or other approach as being the best or more suitable to KO, but claims that both must be reconciled and worked together. The problems that can occur when the ontological approach predominates in a KO research, or in the development of a SOC fail to be addressed by the author, who presents, however, some considerations on the problems of the predominance of epistemological approach.

Supporters of cultural relativism insist that it is better to focus on domain-specific schemes, so as to be aware of their epistemological premises, as any scheme will be biased by the cultural environment in which it has been conceived (Hjørland 2004). However, reference to a general scheme is needed even while indexing special literature (DJ Foskett 1991). If we really want to enable interoperability between different schemes and interdisciplinary research, we will always need some general scheme, at least as a switching device between systems based on different epistemologies. Thus, the need for ontological research in KO is far from being obsolete. (GNOLI, 2008. p. 140).

This statement, in our view, justifies the importance of the two approaches be worked together in KO.

4 Final Considerations

In general, epistemological studies, such as those developed by Hjørland and Gnoli refer to theoretical and methodological currents of an area as paradigms, although based on different classification criteria. We can say that Hjørland's and Gnoli's proposals for the classification of theoretical and methodological currents in KO have been developed from different classification criteria. The first distinguishes paradigms in KO into pragmatic and positivist, and adopts, as criterion, the

epistemological approaches. While the second distinguishes the approaches into ontological and epistemological, and adopts the perspective of object of study in KO as a criterion. We, therefore, consider that this seems to be the main difference between the two proposals: the classification criteria.

Each epistemological approach has its own knowledge validation criteria and uses research methods that aim to meet these criteria. This causes the same research object to be seen differently according to the epistemological stance adopted in the research. With a pragmatic epistemological stance, the object of study is seen as epistemological, and with a positivist epistemological stance, the object of study is seen as ontological. Thus, despite this difference, it appears that these two proposals for classification of theoretical and methodological currents in KO are complementary. For example, the Dewey Decimal Classification (DDC) is classified by Hjørland (2003) as from the historicist paradigm (pragmatic) in KO, and Gnoli (2008), uses DDC and the Universal Decimal Classification (UDC) as examples of the epistemological approach to OC. Hjørland (2003) attributes the analytical-faceted systems to Ranganathan's tradition as from the rationalist paradigm (positivism), because they are supported by rational rules and deductions. Ranganathan developed his classification system based on Henry Evelyn Bliss's works, the Bliss Classification, as well as the Broad System of Ordering (BSO) and the Information Coding Classification (ICC), which are considered by Gnoli (2008) as from the ontological approach .

In addition to these similarities, it appears that both Birger Hjørland as Claudio Gnoli consider it important to integrate approaches to KO. Although Hjørland does not emphasize this aspect in his paper, nor does Gnoli (2008), he makes it clear that the pragmatic approach does not oppose to other epistemological positions; he only considers that isolated evidence are not enough and that the criteria of truth are linked to the objectives and to human activities. (Hjørland, 2003, p. 106). Neither of the two authors propose how this complementarity could occur, and perhaps research in the area can move in this direction, and also in the consolidation of methodologies that incorporate these views.

References

The references was made following the ABNT rules.

FRANCELIN, M. M; KOBASHI, N. Concepções sobre o conceito na organização da informação e do conhecimento. *Ciência da Informação*, Brasília, v. 40 n. 2, 2011.

GNOLI, C. Ten Long-Term Research Questions in Knowledge Organization. *Knowledge Organization*, Würzburg, v.35, n. 2, p.137-149, 2008.

HJØRLAND, B. Concept theory. *JASIST*, New York, v.60, n.8, p.1519-1536, 2009.

HJØRLAND, B. Fundamentals of Knowledge Organization. *Knowledge Organization*, Würzburg, v.30, n.2, p.87-111, 2003.

| 34

HJØRLAND, B. Semantic and Knowledge organization. *ARIST*, v.41, p. 367-405,2007.

PEREGRIN, J. Pragmatism und Semantik. In: Fuhrmann, A.; OLSSON, E. J. *Pragmatisch denken*. p. 89-108. Frankfurt am Main, Germany: Ontos. 2004.

RENDÓN ROJAS, M. A. Hacia um nuevo paradigma em bibliotecologia. *Transinformação*, Campinas, v. 8, n. 3, p. 17-31, set./dez. 1996.

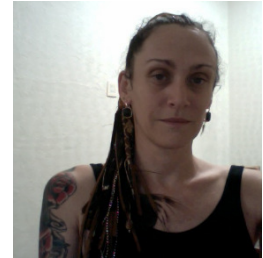
TENNIS, J. T. Epistemology, Theory, and Methodology in Knowledge Organization: Toward a Classification, Metatheory, and Research Framework. *Knowledge Organization*, Würzburg, v. 5, n.2, p. 102-112. 2008.

Historical-epistemological dimension in Knowledge Organization: Gesner's taxonomy contributions, 16th century



Andre Vieira de Freitas Araujo
Federal University of Rio de Janeiro
armarius.araujo@gmail.com

armarius.araujo@gmail.com



Giulia Crippa
University of São Paulo
giuliac@ffclrp.usp.br

| 35

1 Introduction

The historical-epistemological dimension of knowledge organization (KO) is necessary for understanding its nature and its foundations.

There are numerous landmarks that make up the field. Revisiting these landmarks enables constant review of practical and theoretical aspects, as well as the elaboration of new insights and answers to old and new problems that shape the landscape of information and knowledge.

Smiraglia (2013) states that if KO is the science of the order of knowledge, then its scientists should be concerned first with the question of what exactly knowledge is.

Illuminating this perspective, the rescue/retrieval of documentary sources regarding ancient bibliographic practices, whether cataloging and/or classificatory, allows us to expand the understanding of the structure of KO at the present time.

The historical-epistemological dimension of KO is addressed, in this paper, based on the sixteenth century's scheme postulated by Conrad Gesner in *Pandectarum sive partitionum universalium Conradi Gesneri Tigurini, medici & philosophiae professoris, libri XXI: Ad lectores. Secundus hic Bibliothecae nostrae tomus est, totius philosophiae & omnium bonarum artium atque studiorum locos communes & ordines universales simul & particulares complectens.*

From that historical perspective, a general and introductory framework of Gesner's taxonomy and its role as one of the milestones to the organization of knowledge are presented.

2 Conrad Gesner

Conrad Gesner (1516-1565), Swiss scientist, scholar and bibliographer, published books on multiple topics such as linguistics, medicine, theology, botany, zoology, paleontology and mineralogy. He was one of the greatest scientists of the early modern period.

His scientific and scholarly activity,

[...] can be describable as productions of two distinct phases: 1) the formation phase which included classical studies, research in medicine, performing the profession, and the interest in Botany [...] [...] 2) the mature phase, i.e., the scientific production of: a) bibliographical works, b) linguistic-philological works, c) medical, physical and natural sciences works (SABBA, 2012, p 30, our translation).

Gesner's vocation for elaborating bibliographical works, such as the *Catalogus plantarum* (1542), resulted in the most ambitious bibliographical project of Modern Europe: *Bibliotheca Universalis*.

The work was published in four parts, between 1545-1555: 1) *Bibliotheca Universalis, sive Catalogus omnium scriptorum locupletissimus, in tribus linguis, Latina, Graeca, et Hebraica [...]* (1545); 2) *Pandectarum sive partitionum universalium [...]* (1548); 3) *Partitiones theologicae* (1549) and 4) *Appendix bibliothecae* (1555).

According to Malclès (1960), *Bibliotheca Universalis, sive Catalogus [...]* was printed in 1545 by Froschover, in Zurich, on an in-folio of 631 leaves.

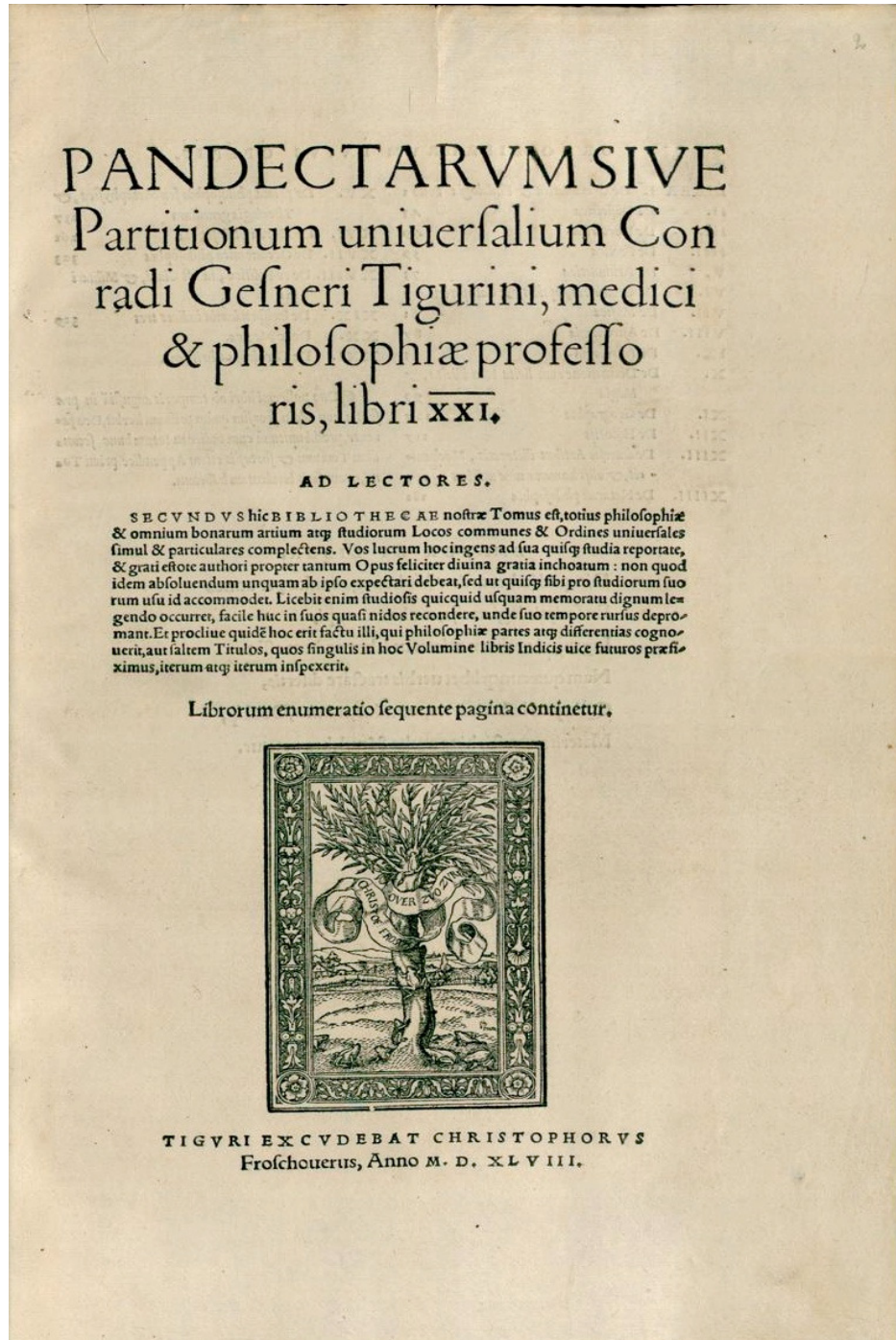
Bibliotheca Universalis, sive Catalogus [...] is an alpha-nominal catalog that presents 5031 authors of works composed in the three classical languages, namely: Latin, Greek and Hebrew. The catalog is organized alphabetically by the first name of authors, but is accompanied by a separate list and organized by surnames (SABBA, 2012). It also presents abstracts and excerpts of the listed documents.

3 *Pandectarum Sive Partitionum Universalium [...]*

Pandectarum sive partitionum universalium [...], also known as *Pandectae* (FIG. 1), was printed in 1548, on an in-folio of 374 leaves (MALCLÈS, 1960). It displays the

documentary material in a classification order, i.e., the presentation of the knowledge extracted from the books of the First Part.

Figure 1 - Pandectae



Source: GESNER (1548).

Available at: <<http://www.e-rara.ch/zuz/content/pageview/624960>>

Initially, the scheme was conceived in XXI Partitions: Pandectae comprise XIX Partitions, so that Partiones theologicae comprises the last Partition, published

separately in 1549 and named Partition XXI. Partition XX, designed to present the semantic catalog of medical works, was not finalized (Serrai,1990).

The back of the title page of Pandectae presents the complete scheme of the general classes of work (FIG. 2), which results in a true repertory of systematic order.

Figure 2 – Complete scheme of Pandectae's general classes

ORDO LIBRORVM HVIVS OPERIS.

LIBER I. De Grammatica & Philologia	Folio 1	XV. De prima philosophia seu Metaphysica, &	
II. De Dialectica	43	Theologia gentilium	337
III. De Rhetorica	49	XVI. De Morali philosophia	261
IIII. De Poetica	59	XVII. De Oeconomica philosophia	303
V. De Arithmetica	73	XVIII. De re Politica, id est Ciuili, & Militari	311
VI. De Geometria, Opticis, & Catoptricis.	77	XIX. De Iurisprudencia indices tres	339
VII. De Musica	81	XX. De re Medica.	
VIII. De Astronomia	87	XXI. De Theologia Christiana.	
IX. De Astrologia	95		
X. De Diuinatione cum licita tum illicita, &			
Magia	99		
XI. De Geographia	107		
XII. De Historijs	117		
XIII. De diuersis Artibus illiteratis, Mechanicis, & alijs humane uitae utilibus	165		
XIIII. De Naturali philosophia	181		

¶ Duo postremi libri ob temporis angustia in praesentia non additi, seorsim quam primum licebit, Deo fauente, prodibunt: una cum Indice in totum hunc secundum Tomum: & fortassis etiam Appendice primi Tomi, quam satis luculentam habemus.

Source: GESNER (1548).

Available at: <<http://www.e-rara.ch/zuz/content/pageview/624961>>

The classification system in Pandectae ranges from the seven liberal arts to the categories of complementary subjects and of interest to scholars and the erudite of the Renaissance.

The scheme is organized as follows: 1-Grammar (and Philology), 2-Dialectics, 3-Rhetorics (representing the trivium), 4-Poetics; 5-Arithmetics, 6-Geometry, 7- Music and 8- Astronomy (the last four classes representing the quadrivium). These are followed by the sciences including the medieval university curriculum: 9 Astrology, 10- Divination and Magic, 11-Geography, 12-History, 13- Mechanical Arts, 14-Natural Philosophy, 15 Metaphysics, 16-Moral Philosophy, 17- Economic philosophy, 18- Politics ; and finally, 19-Law, 20- Medicine and 21- Theology.

Each Pandectae's class represents a book corresponding to a partition. Each book, in turn, is organized as follows: book title (associated with the name of the respective partition), to whom the book is dedicated, and the general list of the subclasses which are subordinated to the class that names the book or partition. Finally, it lists the titles corresponding to each of the subclasses.

Let us examine an example of this system from the De Musica, corresponding to Book VII. On the obverse corresponding to page 81 of Pandectae (Fig. 3), the following information reads:

- The title of the book/partition on the top of the page: "Partitionum universalum Liber VII. De Musica";
- Whom the book is dedicated to: "EXIMIO TYPOGRAPHO BASILIENSI HENRICO PETRO, DE BONIS STVDIIS OPTIME merit, domino & amico suo charissimo, Conradus Gesnerus S.D.P [...]". De Musica is dedicated to Henrich Petri or Henricus Petrus (1508-1579), one of the greatest Master Typographers of Basel who led the Officina Henricpetrina.
- General list of subclasses or sections that make up Book VII: I These are works that address various music genres; II Writings on modal music theory and consonance; III Treaties and introductory materials which address various theoretical and practical aspects of music - Greek, Latin and contemporary; IV Editions of liturgical music; V Editions of secular music, especially Latin, Italian and French; VI Ancient and recent authorities (Gesner's Contemporary) on music; VII Works on musical instruments; VIII Additional books on printed polyphony (appendix) (Bernstein, 1973).

Figure 3 – Partitionum universalum liber VII. De Musica

PARTITIONVM VNIVER

SALIVM LIBER VII. DE MVSICA.

EXIMIO TYPOGRAPHO BASILIENSI HEN-

RICO PETRO, DE BONIS STVDIIIS OPTIME

merito, domino & amico suo charissimo, Contra-
dus Gesnerus S. D. P.



V M fama superioribus annis, quo tempore primum Bibliothecæ Tomum conscribam, incerta ad nos uenisset de clarissimi & doctissimi uiri Henrici Glareani poete laureati Musica, non recte dubius mecum eram, tuò ne Marte & per se hoc argumētū nactasset, an com mentarijs ueterum aliquem illustrasset: utriusq; enim à Glareano, si à quoquā alio, felicissime posse præfari non ignorabam: nunc uero uolumine illo pulcherrimo in lucem ex Officina tua ædico, non ipsius modo Marte conscripum, sed hac in scientia perfectissimum & longe doctissimum esse inel ligo. Quanquam enim ego ipse his de rebus iudicij nihil habeam, eruditorū tamen sententias, quas quidē sæpe consulo, facile & libenter sequor. Propter hoc igitur opus, quo cum ex facundissima Officina tua perpulchre nato, tota fermè ueterum Latinorum & Græcorum Musica renata uidetur, & propter Se uerini Bocij de Musica libros quinq; quos cum omnibus eiusdem scriptis euulgasti, non inepte futurū existimaui ut liber hic de Musica, septimus Partitionum nostrarū, in tuo nomine appareret. Quod autem simpliciter & sine arte hoc facio, tu ipse mihi in causa es. Tanta enim tua in me humanitas est, & multis iam annis fuit, tot tantaq; beneficia tua, ut his pignoribus de familiari proffus & candidissima inter nos amicitia nihil diffusus, *ἀλλὰ πᾶσι τοῖς ἀποστόλοις* animum tibi meum, qualis qualis est, audacter exponā. Hoc interim non tacendū mihi est, quod licet uel Medica uel Theologica uel alia quæuis merito tibi tuo nun cupassem, cum in omni propè scientiarum genere, Latine, Græce, Hebraice & Germanice, tu & felix me moriæ pater tuus, multa præclara uolumina excuderis, non alium tamen præter te cui præsentem hunc libellum æquius inscriberem, reperisse. Proinde quicquid hoc est munusculi, tu tanto humanior libe ratoriq; (nam & in paruis donis non aspernandis liberalitas quæ dā sita est) quanto libellus ipse tenuior, hilari quæso fronte suscipere & aliqua Bibliothecæ nostræ pars esse digneris. Locus hic postulabat cæte ros quodq; omnes uel plurimos & saluberrimos libros, qui ex uberrimo Officinæ tuæ fonte manarunt, nominatim referri, ut pote in omnem posteritatem memorari dignissimos: sed neq; catalogus eorum ad manum mihi nunc est, neq; moram ullam librariorū opera patitur. Dicam tamen ex innumeris pauca, quæ ex tempore mihi occurrunt: In re medica abs te habemus, Alexandrum Trallianum, Constantini Aſri opera omnia, Io. Serapionem (quem Ianum Damascenū Albanus Torinus inscripsit) Symphoriani Campegij præcticam, ut uocant (cuius etiam de omnibus ferè artibus liberalibus & philosophiæ partib; librum ædidiſti) Marcum Gatinariam, Clementinum, Alexandri Benedic̄ti opera omnia medicis certe necessaria, Theophilum de urinis, Philareum de pulsibus, Io. Dantzij dialogos 4. & tabulas de facultati bus simplicium medicamentorum: & alias in Meſuen: & alia non pauca medica quæ iam non memini. Item C. Sollij Sidonij Apollinaris opera, Aurelium Prudentium, T. Lucretiū, Silium Italicum, Statium, Caullum, Tibullum, Propertium, &c. Multos Petri Artopcri, multas deniq; doctissimi uiri Sebastiani Munsteri, uirici tui lucubrationes, quarum postrema Chronicorum uel Geographiæ titulo maximo pas sim applausu excepta est. Generi quoq; tui Hieronymi Curionis opera aliqua, quæ ipse aut parens eius & uiricus excuderunt, hic enumerarem, nisi peculiarem ei de Astronomia librum destinassem. Fax it Christus, à quo uno summa felicitatis nostræ pendet, ut feliciter optimis studijs inferuire, & noua sub inde quæ ad gloriam Dei & fructum studioforum pertineant uolumina publicare pergas. Vale, Tiguri, tertio Calendas Februarij, anno 1548.

TITVLI LIBRI VII. DE MVSICA.

TITVLS I. De Musica in genere.

Italicis, Gallicis.

II. De pertinentibus ad uocem & symphoniam.

VI. De Musicis auctoribus prisctis & nouis.

III. Qui de Musica arte seu methodo tractationes integras aut introductiones, theorice aut practice, scripserint, Græci primum, deinde Latini ueteres & recentiores.

VII. De Musicis instrumentis, & ijs qui eorum usu claruerunt.

III. De cantionibus ecclesiasticis.

APPENDIX cantionum aliquot in diuersis lineis æditarum, quarum nonnullæ etiam ad instrumeta idoneæ sunt: omnes autem ad quartum & quintum Titulum referri possunt.

V. De cantionibus, profanis præcipue, Latinis,

p

Source: GESNER (1548).

Available at: <<http://www.e-rara.ch/zuz/content/pageview/625132>>

Following the first page of Partition De Musica, the list of titles in the scope of each of the subclasses VIII of this partition is presented.

The books listed in Bibliotheca Universalis, sive Catalogus [...] are arranged in the alphabetical order of authors. On the other hand, in The Pandectæ, the books are

ordered according to the loci communes in agreement with their “argomenti”, or the groups of their themes or subjects (Serrai, 1990).

Gesner's scheme reflects his conception on the division of knowledge directly connected with the classical thought of the Renaissance man.

According to Malclès (1960), Gesner's classification system is unique, as it expands the seven liberal arts of the Middle Ages.

When considering the bibliographic details that constituted the books in the sixteenth century, due to the technology of movable types, the Pandectae became the first classification scheme specifically designed to organize books.

Gesner performed a logic re-visitation of sources, while reviewing Johannes Trithemius (1462-1516) whose work *Liber de scriptoribus ecclesiasticis* (1494) is considered as the first bibliographic repertoire of the modern age.

On the other hand, Gesner inspired the work of later bibliographers, as is the case of Florian Treffler, O.S.B. (1483-1565) - librarian and pharmacist of the Benedictine monastery of Benediktbeuren and author of "biblioteconomic manual" *Methodus Exhibens Variorum Indices* (1560).

Gesner organizes each work from previously defined categories and puts his schematic thinking into practice: as he was a naturalistic when working with the classification of beings, he was also a bibliographer when working with the classification of knowledge.

In his scientific and scholarly activity, he was interested in classifying both books and animals. Thus, he promoted a kind of "anatomization" of knowledge, suggesting that the very act of anatomize gave bases for the act of classifying.

4 A brief note on possible contributions of Conrad Gesner to KO

Conrad Gesner's work outlined the bibliographic practices of modern Europe, as many bibliographies were based on *Bibliotheca*. In this sense, Gesner is considered the "Father of Bibliography" and founder of the bibliographic discipline.

Contemporarily, Conrad Gesner is briefly mentioned in papers on descriptive and thematic processing of information, such as in Piedade (1983) on classification; Campello (2006) on bibliographic control; Cunha and Lima (2008) in the field of legal information processing; Mey and Silveira (2009) in the field of cataloging, and Espanha Gomes (2010) in the field of semantic processing of documents.

This scenario can be extended since the organizational thinking in Gesner is unprecedented object of investigation in KO in Brazil, although its role is acknowledged in the foundation of the field.

As Alfredo Serrai (1990, p. 139) reminds us:

Gesner's classification scheme is not simply the map of a series of categorical deviations of bibliographic-literary or bibliographic-disciplinary nature; it is the facade of a concept building, enormously complex, whose internal spaces were modeled from concepts and things ordered according to infallible means of a demonstrable science, supported by reason and sustained by faith (our translation).

The Pandectae are the source and object that construct and evidence, at the same time, an epistemological thread running through the constitution of KO as a field, since on its base lies classification problems that are now clothed in a new guise.

It seems that this is one of the central elements of the work and that it dialogues directly with KO: a reflection on all activities mediating between the production and use of knowledge.

5 Considerations

When Conrad Gesner developed his scheme, he was involved in a process that marked Modern Europe: the expansion of humanists who, interested in the ancient texts and the ways in which they should be organized, dedicated themselves not only to the classification of beings, but also to the classification of knowledge.

What lies on the base of the Pandectae is a dissection stance of the elements of natural sciences which was transferred to the dissection of knowledge.

As we have seen, Gesner's taxonomy was shown innovative in the formation period of Modern Europe, once it expands the seven liberal arts practiced in the Middle Ages, becoming a reference classificatory scheme for being the first scheme specifically designed to organize books.

The challenge that is now posed is to get back to the original ideas of Gesner more deeply, in order to identify the implications for KO, from the historical-epistemological point of view.

This research points to the importance of rescuing Gesner's work not only to expand the comprehension on the underpinnings of KO, but also for the necessary extension of studies and historical debates in this field.

References

The references was made following the ABNT rules.

BERNSTEIN, Lawrence F. The Bibliography of Music in Conrad Gesner's Pandectae (1548). *Acta Musicologica*, v. 45, fasc. 1, jan./jun.1973. pp. 119-163. Disponível em: <<http://www.jstor.org/stable/932225>>. Acesso em: 15 maio 2015.

CAMPELLO, Bernadete. *Introdução ao controle bibliográfico*. Brasília: Briquet de Lemos, 2006.

CUNHA, Murilo Bastos da; LIMA, João Alberto de Oliveira. Tratamento da informação legislativa e jurídica: perspectiva histórica. *Senatus: Cadernos da Secretaria de Informação e Documentação*, Brasília, v. 6, n. 2, p. 33-38, out. 2008.

GESNER, CONRAD. Pandectarum sive partitionum universalium Conradi Gesneri Tigurini, medici & philosophiae professoris, libri XXI : Ad lectores. Secundus hic Bibliothecae nostrae tomus est, totius philosophiae & omnium bonarum artium atque studiorum locos communes & ordines universales simul & particulares complectens [...]. Tiguri : excudebat Christophorus Froschouerus, anno 1548. Disponível em: < <http://www.e-rara.ch/zuz/content/titleinfo/624958>>. Acesso em: 20 fev. 2015.

GOMES, Hagar Espanha. Longa caminhada no tratamento semântico de documentos. In: FREITAS, Lídia Silva de; MARCONDES, Carlos Henrique; RODRIGUES, Ana Célia (Org.). *Documento: gênese e contexto de uso*. Niterói: EDUFF, 2010. p. 193-204.

MALCLÈS, Louise Nöelle. *La bibliografía*. Buenos Aires: EUDEBA, 1960.

MEY, Eliane Serrão Alves; SILVEIRA, Naira Christofolletti. *Catálogo no plural*. Brasília: Briquet de Lemos, 2009.

PIEIDADE, Maria Antonieta Requião. *Introdução à teoria da classificação*. 2.ed. Rio de Janeiro: Interciência, 1983.

SABBA, Fiametta. *La 'Bibliotheca Universalis' di Conrad Gesner*. monumento della cultura europea. Roma: Bulzoni Editore, 2012.

SERRAI, Alfredo. *Conrad Gesner*. Editado por Maria Cochetti. Roma: Bulzoni Editore, 1990.

SMIRAGLIA, Richard. The epistemological dimension of knowledge organization. In: DOBEDEI, Vera Dobedei; GUIMARÃES, José Augusto Chaves (Org.). *Complexidade e organização do conhecimento: desafios de nosso século*. [recurso eletrônico]. Rio de Janeiro : ISKO-Brasil ; Marília : FUNDEPE, 2013. p. 17-25. Disponível em: <<http://isko-brasil.org.br/wp-content/uploads/2013/02/Estudos-avancados-2.pdf>> Acesso em: 20 fev. 2015.

Terminological and interdisciplinary research configurations in knowledge organization (2010-2014)



Leilah Santiago Bufrem
São Paulo State University
santiagobufrem@gmail.com



Aline Elis Arboit
São Paulo State University
aarboit@yahoo.com.br



Juliana Lazzarotto Freitas
São Paulo State University
julilazzarotto@gmail.com

1 Introduction

Studies by Bufrem, Freitas and Nascimento (2014) contribute to outline an overview of the relationship between authors, thematic and theoretical foundations of Knowledge Organization (KO) through citation and co-citation analysis of scientific articles, nationally, and to identify themes and the most representative focus of study. Thus, the structural author characteristics that make up the scientific research scenario in Brazil were visualized.

The possibility to recognize this scenario was based on the conviction that the institutionalization of any area of knowledge occurs through its theoretical and empirical consolidation promoted by the scientific community. Thus, the search for trends concerning the field motivated this study, which is intended to go beyond the mentioned above, deepening knowledge on the field and its interdisciplinary and ideological relations through terms selected by the authors as representative of their

own work regarding "axiological senses" that these terms carry, as pointed by Faraco (2003).

The strong representation of Brazilian authors identified in a previous study reinforced the interest in its production and the noticeable relations from the terms used to represent this production. Therefore, we intend to analyze these terms from the assumption that there are different positions and world views of these authors, embedded in their terminological preferences, and that these axiological positions may vary when analyzed diachronically. The preference for one term over another allows the visualization of interdisciplinary relationships, as the effect of ideological options in KO.

2 Methodological Trajectory

The first stage of this study consisted of a collection and analysis of the scientific production of the most fruitful authors in KO, from national journal articles indexed in the Brazilian Base of Journal Articles in Information Science (Brapci), in the 2003 - 2012 period.

In the second step, we analyzed the terms used as keywords by the authors of the corpus to represent their articles published between 2010 and 2014. As a data source, we used the Lattes Platform, from which we collected, using Script Lattes, these authors' articles in the period. The keywords of each author were identified and organized in Table 1 per year, and the themes represented by colors, which were defined simultaneously to the analytical path, presenting the thematic approaches of the authors to the five major themes related to other fields of knowledge, although the incidence of terms related to KO be identified by color blue. Yellow identifies the keywords of the articles that refer to Epistemology, research methodology, ethics and other philosophic themes; green identifies terms referring to metrical studies and scientific production analysis; red highlights the terms that refer to approaches to education; pink identifies the terms concerning Information Technology; brown is used to highlight terms on terminology, linguistics, semiotics and other Language Studies.

To relate the identified themes and the authors' history of publishing in the past five years, we defined the following parameters: the diachronic distribution of terms and their paradigmatic relations.

The diachronic distribution reveals the regularity of the themes, as it allows to visualize thematic trends in order to analyze them to recover what is considered a significant contribution to the subject in Brazil. Thus, one work with the succession of terms, respecting timelines and processes that define this dimension.

As for the paradigmatic relationship, it reveals consistency among the terms used in the same year, as well as between the other terms of the period for each author. It is conceived that the paradigmatic relationship is established by the concepts of a field, which can express content, methods or values represented by them. Saussure's linguistics defines associative relationships as paradigm to the similar set of elements that are associated in memory and thus form sets (SAUSSURE, 1995).

Moreover, these terms (or set of words) may be considered representative signs of adhesions and choices made by subjects on certain values in the particular context. Borrowing a Bakhtinian concept, it is considered that these terms, given their nature, are loaded with ideological meanings, because everything that is signic is ideological (Bakhtin / Voloshinov, 2010). Such understanding justifies the study of trends or choices made in KO, which, in turn, are juxtaposed to the terms.

3 Analysis and results

Of the eleven most representative authors, only seven have individual authorship studies (26.7%). It is also found in an earlier study by Bufrem, Freitas and Nascimento (2014), that among these authors, six (54.55%) are the head of research groups, while the others are registered as members of these groups. This result alone does not reflect consecration instances, but when one considers the fact that eight of them (72.72%) are CNPq researchers, one can infer that they make up the productive elite in KO in Brazil (BUFREM; FREITAS, NASCIMENTO, 2014).

The universe of authors identified in the exploratory phase was reduced to ten, as one of them did not maintain production in the period between 2010 and 2014. As for the others, 50% have continuous annual production, each from a different university, predominantly those in Southeast and South.

Bufrem, Freitas and Nascimento (2014) highlight the years of 2010 and 2011 as those of higher production. Therefore, it is suggested that the apex of the Brazilian production on KO in these years may be the result of the first ISKO Brazil, held in 2010.

Through the analysis of systematized terms in Chart 1, were found in the production of these authors, the predominance of terms related to KO (in blue), others related Technology (in pink), as in A1 where this theme is preponderant, in addition to a high rate of occurrence of this theme in A4, A8, A10 and A11. It is observed, in A2 and A9, a significant dialogue with metric studies (in green), and epistemological and ethical studies (in yellow). Greater representation of terms related to thematic of language, linguistics, terminology and semiotics (in brown) was detected in A1, A2, A4, A6 and A7. As for educational issues (in red), we noted a more significant occurrence in A2.

| 47

When categorized terms are related to the names of these authors' research groups, there is a convergence not only with the chosen keywords, but also with predominant interdisciplinary relations. This suggests that the produced articles are also results of research, readings and discussions developed within the research groups and, thus, the result of dialogical and social processes contextually located.

Autore s	Research Groups	2010	2011	2012	2013	2014
A1	* Information and knowledge in Cyberspace (Informação e Conhecimento no Ciberespaço)	Search Mechanisms (Mecanismos de busca) Semiotics (Semiótica)	Memory (Memória) Search Mechanisms (Mecanismos de busca) Cyberspace (Ciberespaço)	Search Mechanisms (Mecanismo de busca) Cyberspace (Ciberespaço) Information Technologies (Tecnologias da Informação)	Semiotics (Semiótica) Search Mechanisms (Mecanismos de busca) Intersemiotics Translation (Tradução intersemiótica) Information Technologies (Tecnologias da informação) Scientific Production (Produção científica) Methodological Options (Opções metodológicas) Brappi	Information Architecture (Arquitetura da informação) Public Administration (Administração pública) Semiotic Fold (Dobra semiótica) Gilles Deleuze Invisible Web (Web Invisível) Cyberspace (Ciberespaço) Knowledge Organization (Organização do conhecimento) Scientific Research (Pesquisa científica) Scientific Journal Literature (Literatura periódica científica) Research in Knowledge Organization (Pesquisa em organização do conhecimento) Authorship in Knowledge Organization (Autoria em organização do conhecimento) Citation Analysis (Análise de citação) Coitation Analysis (Análise de coitação) Scientific Production (Produção científica) University Press (Editoras universitárias) Editorial Practices (Práticas editoriais) University Editors (Editores universitários)
A2	* Culture, School Practices and Historic Education (Cultura, Práticas Escolares e Educação Histórica) * Education, Research and Scientific Production (Educação, Pesquisa e Produção Científica)	Scientific Collaboration (Colaboração científica) Citation Analysis (Análise de citações) Epistemology (Epistemologia) Citation Analysis (Análise de citações) Database (Bases de dados) Information Architecture (Arquitetura da informação) Brappi Scientific Journal (Periódico científico) Scientific Communication (Comunicação científica) Education (Educação) Institutional Biography (Biografia institucional) Coauthorship (Coautoria) Information Retrieval (Recuperação da informação) Information Retrieval Method (Método de Recuperação de Informação)	Scientometrics (Cientometria) Scientific Events (Eventos científicos) Scientific Institutionalization (Institucionalização científica) Undergraduate Courses (Cursos de Graduação) Scientific Production (Produção científica) Scientific Methodology (Metodologia científica) Bakhtinian Circle (Círculo de Bakhtin) Dialogism (Dialogismo) Language (Linguagem) Logic (Lógica) Curriculum (Currículo) Teaching (Ensino) Syllabus (Conteúdo disciplinar) Teaching Practice (Prática docente) Didactics (Didática) Mikhail Bakhtin Gaston Bachelard Concept (Conceito) Concept Theory (Teoria do Conceito) Terminology (Terminologia) Conceptual Model (Modelo Conceitual) Social Media (Redes sociais) Bibliometry (Bibliometria)	Scientific Methodology (Metodologia científica) Children's Education (Educação infantil) School Culture (Cultura escolar) Development of Writing (Desenvolvimento da escrita) Knowledge Organization (Organização do Conhecimento) Theoretical Approaches (Aproximações teóricas) Scientific Journal Production (Produção periódica científica) Gender (Gênero) Scientific Journal (Periódico científico) Bibliometry (Bibliometria) Field Education (Educação do campo) History (História) Past (Passado) Historical Significance (Significância histórica) Teacher (Professor) High School (Ensino médio)	Scientific Research (Pesquisa científica) Research Methods (Métodos de pesquisa) University Press (Editoras universitárias) Editorial Practices (Práticas editoriais) Field Education (Educação do campo) School Library (Biblioteca Escolar) Reading (Leitura) Writing (Escrita) Information Science Institutionalization (Institucionalização da Ciência da Informação) Bibliometry (Bibliometria) Citation Studies (Estudos de Citação) GT 7 do ENANCIB Brazil (Brasil) Literary Writing (Escrita Literária) Literary Literacy (Letramento Literário) High School (Ensino Fundamental)	Information Architecture (Arquitetura da informação) Public Administration (Administração pública) Semiotic Fold (Dobra semiótica) Gilles Deleuze Invisible Web (Web Invisível) Cyberspace (Ciberespaço) Knowledge Organization (Organização do conhecimento) Scientific Research (Pesquisa científica) Scientific Journal Literature (Literatura periódica científica) Research in Knowledge Organization (Pesquisa em organização do conhecimento) Authorship in Knowledge Organization (Autoria em organização do conhecimento) Citation Analysis (Análise de citação) Coitation Analysis (Análise de coitação) Scientific Production (Produção científica) University Press (Editoras universitárias) Editorial Practices (Práticas editoriais) University Editors (Editores universitários) Information Science (Ciência da Informação) Scientific Production Analysis (Análise da produção científica) Bibliographic Coupling (Acolamento bibliográfico) Research Methodology (Metodologia da pesquisa) Field Education (Educação do campo) School Library (Biblioteca Escolar) Reading (Leitura) Writing (Escrita) Paraná Teaching Courses (Curso de Pedagogia) PPB Course (Disciplina de PPB) School Disciplines (Disciplinas Escolares) Disciplinary Code (Código Disciplinar) Pedagogic Praxis (Práxis Pedagógica)
A4	* Bioinformatics and Evolutionary Molecular Genetics (Bioinformática e Genética Molecular Evolutiva) * Ontology and Taxonomy: Theoretical and Methodological Aspects (Ontologia e Taxonomia: Aspectos Teóricos e Metodológicos) * Conceptual Modeling for Hypertextual Organization of Documents (Modelagem Conceitual para Organização Hipertextual de Documentos – mhtx).	Terminology (Terminologia) Onomasiological Approach (Abordagem Onomasiológica) Semasiological Approach (Abordagem Semasiológica) Definition (Definição) Semantics Compatibility (Compatibilização semântica)	Information Extraction (Extração de informação) Ontology Reuse (Reuso de ontologia) Ontology Alignment (Alinhamento de ontologia)	-	Knowledge Organization (Organização do conhecimento) Representation Theory (Teorias de Representação) Categorization (Categorização) Domain Analysis (Análise de Domínio) Genomics Annotation (Anotação Genômica) Ontology Reuse (Reuso de ontologia) Ontology Alignment (Alinhamento de ontologia) Language Compatibility (Compatibilidade de linguagem) Trypanosomatids (Trypanosomatídeos) Knowledge Acquisition (Aquisição de conhecimento)	--
A5	* RESNAPAP * TEMMA * SCIENTIA		Scientific Institutionalization (Institucionalização científica) Undergraduate Courses (Cursos de Graduação) Tourism (Turismo)		Bibliometry (Bibliometria) Scientometrics (Cientometria) ICT Indicators (Indicadores de CT&I) Information Visualization (Visualização de Informação)	

			Brazil (Brasil)			
--	--	--	-----------------	--	--	--

4 Final Considerations

The terms diachronic analysis revealed predominant domains that interact with KO, indicating that these domains also affect, in a balanced manner, over the years, and that KO does not prioritize one over another at specific times. This fact denotes the constant interdisciplinary work among authors of the area and the variety of approaches and potential research fronts to the field.

Thus, the validity of this study resided in the possibility of joint empirical research with theoretical and interdisciplinary framework that can be viewed from the established relationships. Among them, the agreement of the keywords used and the themes studied in the authors' research groups indicate an approximation of the authors' scientific production to their lines and groups. It is remarkable that the associative relationships between the terms of the same authors stand out as an indication of coherence in this scientific production and of deepening of the themes studied over the years. It is suggested, therefore, that this approximation can be considered a strong point in the production representing these authors, considering that the legitimate participation in research groups presumes possibilities for deepening and continuing research in co-authorship with members of the same group.

References

The references was made following the ABNT rules.

BAKHTIN, M. (VOLOCHÍNOV). *Marxismo e filosofia da linguagem*. 14. ed. São Paulo: Hucitec, 2010.

BUFREM, L. S.; FREITAS, J. L.; NASCIMENTO, B. S. Autoria e pesquisa em Organização do Conhecimento: análise da produção científica em Ciência da Informação. *Revista da Faculdade de Biblioteconomia e Comunicação da UFRGS*, Porto Alegre, v. 20, n. 3, ed. especial, 2014.

FARACO, C. A. *Linguagem e diálogo: as ideias lingüísticas do círculo de Bakhtin*. Curitiba: Criar, 2003.

FREITAS, J. L.; BUFREM, L. S.; NASCIMENTO, B. S. A organização do conhecimento na dinâmica da pesquisa em artigos da literatura científica da Brapci. *Transinformação*, Campinas, v. 26, n. 3, 2014.

SAUSSURE, F. *Curso de lingüística geral*. São Paulo: Cultrix, 1995.

The conceptual dimension of Knowledge Organization in NASKO conferences: Bardian's content analysis



José Augusto Chaves Guimarães
São Paulo State University
guima@marilia.unesp.br



André Ynada dos Santos
São Paulo State University
andre.ynada@gmail.com



Rodrigo de Sales
Federal Fluminense University
rodrigo.sales.s@gmail.com



Daniela Fernanda de Oliveira Matos
São Paulo State University
danielamatos@marilia.unesp.br

1 Introduction

Knowledge Organization plays a key role in Information Science, with theoretical advance markedly from the creation of ISKO in 1989. However, it is observed that this field is still in consolidation, in search of construction and delimitation its own conceptual basis and limits, due to different theoretical influences. Such aspect is revealed, among others, through a diversity with respect to its conceptual aspect, particularly in relation to its object, which leads to the need to examine how ISKO's scientific environment has been building/ delimiting this conceptual dimension, from its official scientific discourse internationally.

Thus, research conducted by Guimarães et al (2014) regarding the universe of ISKO's international congresses demonstrated that knowledge organization is an area or

field of knowledge and/or application related to the construction of specialized discourses. Its operational nature lies either aimed at the organization and representation of concepts, or focused on the search for information and access to knowledge, particularly in dialogue with information retrieval.

Having as object the structure of knowledge contained in documents (recorded, socialized and published knowledge), with special emphasis on the concepts and modeling as well as the discursive activities in scientific domains and specific social and cultural practices, knowledge organization uses instruments such as classification systems, thesauri, vocabulary and other indexing languages. In this context, the predominant approaches are related to cognitive, technological, logical and conceptual, socio-cultural and management issues, pointing to the most urgent dialogues in the area.

In this opportunity, and continuing the mentioned research, we analyze the conceptual issue of knowledge organization in congresses held by the North American chapter of ISKO.

In this view, we part from a brief approach to knowledge organization as a field of study to get to the context of International Society for Knowledge Organization - ISKO, either in its international organization, or in its national or regional chapters.

Next, we describe the data collection procedure, from the proceedings of the North American chapters of ISKO, which are presented, analyzed and discussed under two of the eleven domain analysis approaches proposed by Hjørland (2002): a) epistemological, from the content analysis of presented definitions, and b) bibliometric, from citation networks that support this production.

In the context of information science, knowledge organization and representation presents a mediating nature, configuring a set of processes that establish the mediation between knowledge, which, once produced, was materialized and socialized, so that this knowledge can serve as a basis for generating new knowledge. This new knowledge, in turn, once materialized and socialized, may also be the object of new organization and representation, characterizing what can be named as the helical flow of information (GUIMARÃES, 2008).

Thus, as Garcia Marco (1997, p.8) recalls, knowledge organization emerges as a branch of studies most notably from the 1990s,

[...] at the crossroads of so-called Cognitive Sciences, in the field meeting Human Knowledge Sciences (Neuroscience, Psychology and Epistemology), Information Sciences and Communication (including Semiotics and Linguistics), Mathematics (including Logic and formal Languages) and Computer Science.

The term Knowledge Organization was introduced by Henry Evelyn Bliss, from his works "The organization of Knowledge and systems of sciences", 1929 and "The organization of knowledge in libraries", 1933. Referring to the pioneer work of Evelyn Bliss, (The organization of knowledge and the system of sciences), Ingetraut Dahlberg (1993, p.211) states that:

we are dealing here with a volume of knowledge collected, deepened and matured over many centuries which, however, only today has been recognized as being an autonomous Field of knowledge which needs to find its proper place in society na which demands to be recognized as such in the system of the sciences.

Later the term was rescued and discussed by Dagobert Soergel (1971) and Ingetraut Dalhberg (1993).

To Dahlberg (2008), Knowledge Organization has a scientific nature and aims to systematize knowledge units (concepts) from the elements or characteristics inherent to them. To this, application of concepts and classes of concepts that can allow the transfer of content (subjects) is added. For the author (Dalhberg, 1993) two knowledge concepts should be considered: as an individual process, which is not transferable, and the one that presupposes a consensus, from a record, the latter object of knowledge organization. Thus, the term Knowledge Organization for Information Science will be dedicated to activities to classify, index and represent knowledge through records with the specific purpose of meeting the need for information (BUFREM, 2008).

In the research context, Knowledge Organization has dedicated to the study of laws, principles and procedures on which knowledge is supported, whose context highlights the work of the International Society for Knowledge Organization - ISKO, established in 1989 in Germany, by Ingetraut Dahlberg's initiative as a society of researchers addressing the theme of Knowledge Organization. Its mission consists in promoting theoretical and applied advances of Knowledge Organization in different fields

and in different manners, bringing together researchers from different countries, in areas such as Information Science, Linguistics, Philosophy, and Computer Science, among others.

ISKO's scientific activities include international conferences, which take place in even years, national or regional conferences in odd years, and two scientific publications⁵⁴ of major impact in the field: the journal Knowledge Organization, and the series Advances in Knowledge Organization.

High point of the scientific activities of ISKO, the international conferences have been taking place every two years as follows: Darmstadt, 1990; New Delhi, 1992; Copenhagen, 1994; Washington, 1996; Lille 1998; Toronto, 2000; Granada, 2002; London, 2004; Vienna, 2006; Montréal, 2008; Rome, 2010; Mysore, 2012; and Krakow, in 2014, and Rio de Janeiro, 2016.

In its international relations, ISKO maintains collaborative activities with UNESCO, the European Commission, ISO (International Standardization Organization), IFLA (International Federation of Library Associations), ASIST (American Society for Information Science and Technology), among others .

Operationally, ISKO develops its activities from national or regional chapters, which, in turn, conduct biennial congresses. Today, the following chapters are: Germanic, North American, French, Spanish, Italian, Brazilian, Polish, Scandinavian, Maghreb and India.

North America ISKO - NASKO, established in 2007, is formed by Canada and the United States, having held four conferences on the themes: Knowledge Organization (Toronto, 2007), Pioneering North American Contributions to Knowledge Organization (Syracuse, 2009); Expanding Our Horizons, Evaluating Our Parameters (Toronto, 2011) and Cultures, Transition KO: Evolving Exploration, Critical Reflection, and Practical Work (Milwaukee, 2013).

2 Methodology

This research was developed from the set of proceedings of the North American Chapter meetings (2007, 2009, 2011 and 2013) of ISKO (International Society for Knowledge Organization). The choice for those chapters was because although a recent

chapter, it proves to be very productive in the area internationally, and with a strong presence in international congresses of ISKO and in the journal Knowledge Organization.

The research focused on the approach of domain analysis which, in the international context of Information Science, has been traditionally worked in theoretical and applied terms by Hjørland & Albrechtsen (1995); Moya Anegón & Herrero Solana⁵⁵ (2001), Hjørland (2002, 2004), Tennis (2003) and Smiraglia (2011), among others. It constitutes an important approach for characterization and evaluation of science, as it allows to identify conditions by which scientific knowledge is constructed and socialized. Therefore, domain analysis, according to Smiraglia (2012, p.114) "... a group with an ontological base that reveals an underlying teleology, a set of common hypotheses, epistemological consensus on methodological approaches, and social semantics "

Thus, we were based on two of the eleven domain analysis approaches proposed by Hjørland (2002): the epistemological and bibliometric.

Therefore, and following the methodological approach used by Guimarães, Martínez Ávila, Sales and Alencar (2014) we carried out a research on the complete collection of the proceedings, based on the incidence of the term "knowledge organization" (and its correspondents in French, Spanish and Portuguese), variations in French, in the titles, keywords, abstracts and headings of section of the published communications.

Next, the reading of the communications that included such incidence was performed, by selecting those that brought concepts, definitions and theoretical considerations about the nature of knowledge organization.

The epistemological dimension was approached from the conceptual universe of the analyzed literature. Thus, regarding the definitions, we proceeded to Content Analysis (Bardin, 2003), whose choice is justified by the fact that its procedures allowed an analysis based on inferences drawn from the content of documents - from an interpretation controlled through variables or indicators that provided greater freedom to the analyst, without losing objectivity and because it is an analysis guided by definitions conceived by authors/researchers, i.e., arguments textually recorded.

The analysis was divided into three phases: pre-analysis, exploration of the material and treatment of the results (inferences and interpretations). Therefore, the

delimitation of the corpus concerning the four basic rules proposed by Bardin (2003): completeness (all definitions found in data collection were considered useful material to the analysis, with no exclusion process); representativeness (ensured by the choice of scientific communication channel as the papers published in the proceedings of ISKO are authored by researchers and theorists representing a significant share of the international production in knowledge organization); homogeneity (the analyzed definitions are related to the same theme, thus, they are homogeneous both in textual structure and in subject); and relevance of the material (the sources are the ideas explicated by the authors/researchers in their definitions).

The definitions were analyzed from the following categories: Nature; Object, Objective, Instruments, Processes, Interdisciplinary Dialogues and Perspectives or Approaches.

Following, and in order to characterize the domain, we carried out a bibliometric analysis, specifically in relation to citation analysis, from the sources (theoretical references) used in the selected papers, in order to identify dialogues, consonance in theoretical currents etc.

Through citation analysis, it becomes possible to analyze the impact and visibility of authors in a given scientific domain, being particularly useful to identify theoretical schools. Thus, it constitutes a special tool for, from qualitative and quantitative analysis, better understand the epistemological universe of a given domain.

From these studies it is possible to construct and graphically visualize social citation networks, especially useful to visualize the behavior of a given knowledge domain. For visualization purposes of the citation networks, we used the software PAJEK described by Adamic (2008).

3 Result presentation, analysis and discussion

We parted from a set of 66 papers published in NASKO. From this set, and from the incidence of the terms in the title, we retrieved a corpus of 51 studies (Abbas, 2007; Adler & Tennis, 2013; Beak et al, 2013; Bernstein, 2009; Buchel & Hill, 2009; Campbell, 2009, 2011; Campbell et al., 2007; Dousa, 2009a, 2009b, 2011; Edwards, 2011, 2013; Feinberg,

2007; Fox, 2011; Fox & Reece, 2013; Frické, 2013; Graf & Smiraglia, 2013; Green, 2009, 2011; Green & Fallgren, 2007; Guimarães & Sales, 2013; Heuvel, 2011; Hoffman, 2009; Hudon, 2011; Hudon & Guitard, 2013; Jacob & Zhang, 2013; Kasten, 2007; Kemp, 2007; La Barre, 2009; Loehrlein, 2011; MacCall, 2011; Marchese & Smiraglia, 2013; Martínez-Ávila & Smiraglia, 2013; McTavish & Fortier, 2011; Milani & Guimarães, 2011; Moulaisoh;⁵⁷ Dykas & Budd 2013; Olson & Howarth, 2013; Pattuelli, 2011; Pimentel, 2007, 2009, 2011; Schatz; Stennes-Spidahl; Mills & Loehrlein, 2013; Smiraglia, 2007, 2009, 2013; Szostak, 2013; Tennis, 2007; Thornton, 2011; Tognoli; Guimarães & Tennis, 2013; and Zhang, Kipp & Lee, 2013).

Assuming that an analysis of terminological character does not seek unanimity, but concepts that stand out on a corpus referring to an analyzed subject, it is possible to highlight, according to most of the texts analyzed, that knowledge organization has been predominantly addressed by NASKO researchers from a systematic instrumental perspective, as the incidence of approach to KO as a system, or a systematic form of knowledge structures is the most frequent approach in the investigated texts. It is also noted that, in the universe analyzed, in 26 papers, the fact that knowledge organization is linked to the issue of knowledge organization systems is highlighted. This fact does not necessarily mean that the authors understand KO merely as a system, but rather, they address the aspects of knowledge organization highlighting its systematic character of knowledge structuring.

It is necessary to consider the fact that the variables presented above are not mutually exclusive, for a detailed observation shows that many authors use different perspectives about KO, such as text written by Olson & Howarth (2013), for example. This text addresses KO under various natures: as a system, as a specialized subject, as a process, or even as a community. The incidence of texts that address KO from different perspectives is more frequent than expected, and it can reveal, among other aspects, at least two facts: the multifaceted nature of knowledge organization and the incipient discussion of its nature.

The term knowledge organization was linked to knowledge organization systems in the following studies: Abbas, 2007; Feinberg 2007; Green & Fallgren, 2007; Pimentel, 2007; Kasten, 2007; Bernstein, 2009; Buchel & Hill, 2009; Pimentel, 2009; Green, 2009;

Campbell, 2009; Green, 2011; Heuvel, 2011; Edwards, 2011; Pattuelli, 2011; Pimentel, 2011; Fox, 2011; McTavish & Fortier, 2011; Martinez-Avila & Smiraglia, 2013; Moulaison, Dykas & Budd 2013; Graf & Smiraglia, 2013; Adler & Tennis, 2013; Fox & Reece, 2013; Frické, 2013; Guimaraes & Sales, 2013; Olson & Howarth, 2013; Schatz; Stennes-Spidahl, Mills & Loehrlein, 2013.

Knowledge organization understood as a process is described in the following papers: Campbell et al, 2007; Kasten, 2007; Pimentel, 2007; Kemp, 2007; Tennis, 2007; Hoffman, 2009; Fox, 2011; MacCall, 2011; Milani & Guimarães, 2011; Adler & Tennis, 2013; Fox & Reece, 2013; Edwards, 2013; Jacob & Zhang, 2013; Olson & Howarth, 2013.

However, knowledge organization addressed as an investigative field is presented in the papers: Feinberg 2007; Kasten, 2007; Smiraglia, 2007; Smiraglia, 2009; La Barre, 2009; Bernstein, 2009; Campbell, 2011; Loehrlein, 2011; Milani & Guimarães, 2011; Zhang, Kipp & Lee, 2013; Adler & Tennis, 2013; Marchese & Smiraglia, 2013; Tognoli; Guimarães & Tennis, 2013.

Next, we verify knowledge organization addressed as: a) community: Pimentel, 2007; Campbell, 2009; Smiraglia, 2009; Edwards, 2013; Moulaison, Dykas & Budd 2013; Thornton, 2013; Olson & Howarth, 2013; b) as a domain: Smiraglia, 2009; Beak et al, 2013; Graf & Smiraglia, 2013; Smiraglia, 2013; Guimaraes & Sales, 2013 or; c) as a theme of expertise: Dousa, 2009; Hudon, 2011; Hudon & Guitard, 2013; Olson & Howarth, 2013.

Interpretations and inferences reveal that, although there is a predominance in linking KO to knowledge organization systems, the multi facets is a hallmark in the conceptualization of KO under NASKO.

Following the KO approach as a system (present in 26 texts), the most present perspectives were knowledge organization as a process (or set of practices and procedures), present in 14 texts, and as field of study, presented in at least 13 analyzed texts. Then the most popular perspectives in the texts were of KO as a scientific community (7 texts), while as a domain (in 5 texts) and as a specialized subject (4 texts).

The study included 593 cited authors (excluding collective entities and self-citations), and authors who received two or more citations: 159 (18.5%), from a total of 752 citations.

Applying Price Theory of Elitism, we got to a group of 27 authors, and one author with 5 citations. Thus, we considered all authors cited at least five times, which led to a group of 29 authors.

Based on these aspects, we constructed citation networks using Pajek software, as can be seen below:

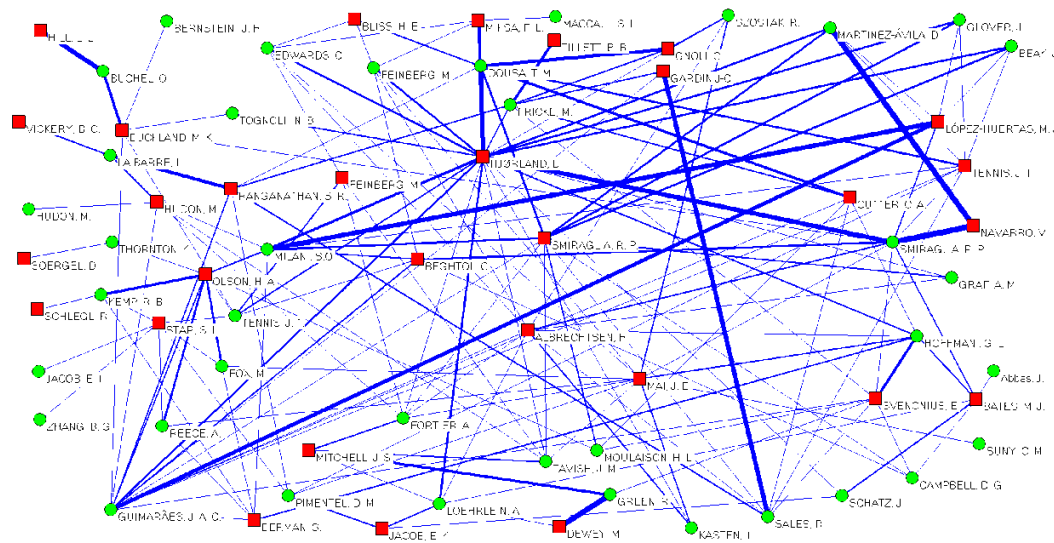


Figure 1: Citation Network. Source: by the authors

Caption: Green circles - citing authors; Red Squares - cited authors.

Among the cited authors, we highlight the number of researchers who cite B. Hjørland, R.P. Smiraglia, H.A. Olson, M.J. López-Huertas and V. Navarro, evidenced by the large number of citing vectors in relation to those cited.

The strongest links between citing researchers and cited authors are observed in J.A.C. Guimarães, S.O. Milani to M.J. López-Huertas, R. Sales for J-C. Gardin and D. Martinez-Avila, R.P. Smiraglia to V. Navarro, evidencing the strong link of this citing community in relation to these authors. In addition, we highlight the intense citation of Hjørland by J.A.C. Guimarães, S.O. Milani and R.P. Smiraglia.

4 Conclusion

If, on the one hand, the analysis results do not allow more definitive assertions, because the variables are not mutually exclusive, impeding to establish, how the North American KO community conceives the nature of knowledge organization, on the other hand, it can be inferred that whatever the perspective adopted to conceptualize KO, their connection with knowledge organization systems or knowledge systematization efforts is strongly addressed by North American researchers. ⁶⁰

Traditionally, the North American current leveraged, in the late nineteenth and early twentieth century, the scientific discourse of standardization of data processing procedures for the construction of products and services for libraries, which strongly echoed in the international arena. It can be said that, in a way, this is also reflected in this new space named KO, as the systems and processes are still dominant in the literature. However, it is not hard to see that this more pragmatic perspective has been, for some time, sharing space within NASKO, with more pronounced epistemological concerns, giving rise to perspectives which prefer to address knowledge organization as field, domain or scientific community.

References

The references was made following the ABNT rules.

ABBAS, J. In the margins: Reflections on scribbles, knowledge organization, and access (extended abstract). In: *North American Symposium on Knowledge Organization, 1, 2007*, Toronto, *Proceedings of the North American Symposium on Knowledge Organization*, p. 107-111. Available: <http://iskocus.org/nasko2007-proceedings.php>.

ADAMIC, L. *Introductory social network analysis with Pajek*. University of Michigan, 2008. Available at: http://ocw.mit.edu/courses/economics/14-15j-networks-fall-2009/assignments/MIT14_15JF09_pajek.pdf. Access at 14.07.2014.

ADLER, M.; TENNIS, J. Toward a taxonomy of harm. In: *North American Symposium on Knowledge Organization, 4*, Milwaukee, 2013, *Proceedings of the North American Symposium on Knowledge Organization*, p. 1-19.

ANDERSEN, J. Communication technologies and the concept of knowledge organization. *Knowledge Organization*, Wuzburg, v. 29, p. 29-39, 2002.

BARDIN, L. *L'analyse du contenu*. 7. ed. Paris: PUF, 2003. 296 p.

BEAK, J., GLOVER, J., MARTÍNEZ-ÁVILA, D., MILANI, S.O. International Comparative Domain Analysis in Knowledge Organization Research Topics in Four Countries - Brazil, South Korea,

Spain and the United States. In: *North American Symposium on Knowledge Organization, 4*, Milwaukee, 2013, *Proceedings of the North American Symposium on Knowledge Organization*, Available: <http://iskocus.org/nasko2013-proceedings.php>.

BEAK, J.; OLSON, H. A. Analysis of metadata schemas for children's libraries. In: *North American Symposium on Knowledge Organization, 3*, Toronto, 2011, *Proceedings of the North American Symposium on Knowledge Organization*, p. 13-22. Available: <http://iskocus.org/nasko2011-proceedings.php>.

1

BERNSTEIN, J. H. The data-information-knowledge-wisdom hierarchy and its antithesis. In: *North American Symposium on Knowledge Organization, 2*, Syracuse, 2009, *Proceedings of the North American Symposium on Knowledge Organization*, p. 68-75. Available: <http://iskocus.org/nasko2009-proceedings.php>.

BUCHER, O.; HILL, L. Treatment of Georeferencing. in Knowledge Organization Systems: North American Contributions to Integrated Georeferencing. In: *North American Symposium on Knowledge Organization, 2*, Syracuse, 2009, *Proceedings of the North American Symposium on Knowledge Organization*, p. 47-57. Available: <http://iskocus.org/nasko2009-proceedings.php>.

BUFREM, L.S. Práticas de organização e divulgação da produção intelectual em ciência da informação no Brasil. *Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação*, Florianópolis: v. 13, n. esp., p. 36-53, 1º sem. 2008.

CAMPBELL, D.G. RDA and RDF: A discourse analysis of two standards of resource description. In: *North American Symposium on Knowledge Organization, 3*, Toronto, 2011, *Proceedings of the North American Symposium on Knowledge Organization*, p. 207-216. Available: <http://iskocus.org/nasko2011-proceedings.php>.

CAMPBELL, D.G. Tensions between language and discourse in north american knowledge organization. In: *North American Symposium on Knowledge Organization, 2*, Syracuse, 2009, *Proceedings of the North American Symposium on Knowledge Organization*, p.10-16. Available: <http://iskocus.org/nasko2009-proceedings.php>.

CAMPBELL, D.G.; BRUNDIN, M.; MACLEAN, G.; BAIRD, C. Everything old is new again: Finding a place for knowledge structures in a satisficing world. In: *North American Symposium on Knowledge Organization, 1*, 2007, Toronto, *Proceedings of the North American Symposium on Knowledge Organization*, p. 21-30. Available: <http://iskocus.org/nasko2007-proceedings.php>.

DAHLBERG, I. Knowledge organization: its scope and possibilities. *Knowledge Organization*, Frankfurt, v.20, n.4, p.211-222, 1993.

DAHLBERG, I. Knowledge organization: a new science? *Knowledge organization*, Wuzburg, v. 33, p. 11-2, 2006.

DAHLBERG, I. Teoria do conceito. *Ciência da Informação*, Rio de Janeiro, v.7, n. 2, p. 101-107, 1978.

DAHLBERG, I. The Information Coding Classification (ICC): A modern theory-based fully-faceted, universal system of knowledge field. *Axiomathes*, v.18, n.2, p.161-176, 2008.

DOUSA, T.M. Classical Pragmatism and its Varieties: On a Pluriform Metatheoretical Perspective for Knowledge Organization. In: *North American Symposium on Knowledge Organization, 2*, Syracuse, 2009a, *Proceedings of the North American Symposium on Knowledge Organization*, p.1-9. Available: <http://iskocus.org/nasko2009-proceedings.php>.

DOUSA, T.M. Evolutionary Order in the Classification Theories of C. A. Cutter & E. C. Richardson: Its Nature and Limits. In: *North American Symposium on Knowledge Organization, 2*, Syracuse, 2009b, *Proceedings of the North American Symposium on Knowledge Organization*, p. 76-90. Available: <http://iskocus.org/nasko2009-proceedings.php>.

i2

DOUSA, T.M. Concretes, countries, and processes in Julius O. Kaiser's Theory of Systematic Indexing: A case study in the definition of general categories. In: *North American Symposium on Knowledge Organization, 3*, Toronto, 2011, *Proceedings of the North American Symposium on Knowledge Organization*, p. 160-173. Available: <http://iskocus.org/nasko2011-proceedings.php>.

EDWARDS, C. Soviet classifications and situated knowledges. In: *North American Symposium on Knowledge Organization, 3*, Toronto, 2011, *Proceedings of the North American Symposium on Knowledge Organization*, p. 13-22. Available: <http://iskocus.org/nasko2011-proceedings.php>.

EDWARDS, C. The Social Role of Public Library Classifications. In: *North American Symposium on Knowledge Organization, 4*, Milwaukee, 2013, *Proceedings of the North American Symposium on Knowledge Organization*. Available: <http://iskocus.org/nasko2013-proceedings.php>.

FEINBERG, M. Beyond retrieval: A proposal to expand the design space of classification. In: *North American Symposium on Knowledge Organization, 1*, 2007, Toronto, *Proceedings of the North American Symposium on Knowledge Organization*, p. 31-43. Available: <http://iskocus.org/nasko2007-proceedings.php>.

FOX, M. Prototype theory: an alternative concept theory for categorizing sex and gender. In: *North American Symposium on Knowledge Organization, 3*, Toronto, 2011, *Proceedings of the North American Symposium on Knowledge Organization*, p. 13-22.

FOX, M.; REECE, A. The impossible decision: social tagging and Derridas' s deconstructed hospitality. In: *North American Symposium on Knowledge Organization, 4*, Milwaukee, 2013, *Proceedings of the North American Symposium on Knowledge Organization*, p. 58-74. Available: <http://iskocus.org/nasko2013-proceedings.php>.

FRICKÉ, M. Logic and the organization of information: an introduction. In: *North American Symposium on Knowledge Organization, 4*, Milwaukee, 2013, *Proceedings of the North American Symposium on Knowledge Organization*, p. 75-80. Available: <http://iskocus.org/nasko2013-proceedings.php>.

FURNER, J. Interrogating "identity": a philosophical approach to an enduring issue in knowledge organization. *Knowledge organization*, Wuzburg, v. 36, p. 3-16, 2009.

GARCIA MARCO, F.J. Avances en Organización del Conocimiento en España. In: Garcia Marco, F.J. (ed.) *Organización del conocimiento en sistemas de información y documentación*. Getafe: Libr. General, 1997.

GILCHRIST, A. A note on the theory and practice of knowledge organization and knowledge representation. *Scire*, Zaragoza, v. 17, n. 1, p. 23-28, 2011.

| 63

GMÜR, M. Co-Citation Analysis and the Search for Invisible Colleges: A Methodological Evaluation. *Scientometrics*, v. 57, n. 1, p. 27-57, 2003.

GNOLI, C. Ten long-term research questions in knowledge organization. *Knowledge organization*, Wuzburg, v. 35, p. 137-49, 2008.

GNOLI, C. Ontological foundations in KO: the theory of integrative levels applied in citation order. *Scire*, Zaragoza, n. 17, v. 1, p. 29-34, 2011.

GRAF, A.; SMIRAGLIA, R.P. Transition in Education: Domain Analysis from the Encyclopedia of Milwaukee. In: *North American Symposium on Knowledge Organization*, 4, Milwaukee, 2013, *Proceedings of the North American Symposium on Knowledge Organization*. Available: <http://iskocus.org/nasko2013-proceedings.php>.

GREEN, R. Melvil Dewey's Ingenious Notational System. In: *North American Symposium on Knowledge Organization*, 2, Syracuse, 2009, *Proceedings of the North American Symposium on Knowledge Organization*, p. 91-99. Available: <http://iskocus.org/nasko2009-proceedings.php>.

GREEN, R.; FALLGREN, N. Anticipating new media: A faceted classification of material types. In: *North American Symposium on Knowledge Organization*, 1, 2007, Toronto, *Proceedings of the North American Symposium on Knowledge Organization*, p. 87-99. Available: <http://iskocus.org/nasko2007-proceedings.php>.

GREEN, R. See-also relationships in the Dewey Decimal Classification. In: *North American Symposium on Knowledge Organization*, 3, Toronto, 2011, *Proceedings of the North American Symposium on Knowledge Organization*, p. 41-51. Available: <http://iskocus.org/nasko2011-proceedings.php>.

GUIMARÃES, J.A.C. *A dimensão conceitual da organização do conhecimento no universo científico da Internacional Society for Knowledge Organization-ISKO: uma análise de domínio com subsídio da garantia literária*. Marília: UNESP, 2011. Projeto de Produtividade em Pesquisa (PQ)- CNPq referente ao período 2012-2016.

GUIMARÃES, J.A.C. A dimensão teórica do tratamento temático da informação e suas interlocuções com o universo científico da International Society for Knowledge Organization (ISKO). *Revista Ibero-Americana de Ciência da Informação*, v. 1, n. 1, 2008.

GUIMARÃES, J.A.C. ; SALES, Rodrigo de ; MARTINEZ-ÁVILA, D.; ALENCAR, M. F. The conceptual dimension of knowledge organization in the ISKO proceedings domain: a Bardinian analysis. In: Wieslaw Babik. (Org.). *Knowledge organization in the 21st century: between historical patterns and future perspectives*. Wurzburg: Ergon Verlag: 2014, p. 101-106.

GUIMARÃES, J.A.C.; SALES, R. Contributions of Julius Otto Kaiser to the theoretical framework of Knowledge Organization: basis for the analytical-synthetic method. In: *North American Symposium on Knowledge Organization, 4*, Milwaukee, 2013, *Proceedings of the North American Symposium on Knowledge Organization*. Available: <http://iskocus.org/nasko2013-proceedings.php>.

HJØRLAND, B. Domain analysis: a socio-cognitive orientation for Information Science research. *Bulletin of the American Society for Information Science and Technology*, [S.l.]: v. 30, n.3, feb./mar. 2004. Available at <<http://www.asis.org/Bulletin/Feb-04/hjorland.html>>. p64

HJØRLAND, B. What is knowledge organization (KO)? *Knowledge organization*, Wurzburg, v. 35, n. 2/3, p. 86-101, 2008.

HJØRLAND, B. Concept theory. *JASIST*, New York, v. 60, n. 8, p.1519-36, 2009.

HJØRLAND, B. Domain analysis in information Science: eleven approaches traditional as well as innovative. *Journal of Documentation*, London, v. 58, n. 4, p. 422-462, 2002.

HJØRLAND, B. Meta-analysis should also be visible inside information science (Letter to the Editor). *Journal of the American Society for Information Science and Technology*, New York, v.53, n.4.

HJØRLAND, B.; ALBRECHTSEN, H. Toward a new horizon in information science: domain-analysis. *Journal of the American Society for Information Science*, New York, v. 46, n. 6, p.400-425, 1995.

HOFFMAN, G.L. Applying the User-Centered Paradigm to Cataloging Standards in Theory and Practice: Problems and Prospects. In: *North American Symposium on Knowledge Organization, 2*, Syracuse, 2009, *Proceedings of the North American Symposium on Knowledge Organization*, p. 27-34. Available: <http://iskocus.org/nasko2009-proceedings.php>.

HUDON, M. Teaching bibliographic classification In The 21st Century. In: *North American Symposium on Knowledge Organization, 3*, Toronto, 2011, *Proceedings of the North American Symposium on Knowledge Organization*, p. 63-73. Available: <http://iskocus.org/nasko2011-proceedings.php>.

HUDON, M. KO and Classification Education Objectives: Are we keeping up with the transformation of our field? In: *North American Symposium on Knowledge Organization, 4*, Milwaukee, 2013, *Proceedings of the North American Symposium on Knowledge Organization*. Available: <http://iskocus.org/nasko2013-proceedings.php>.

JACOB, E. K.; ZHANG, G. The role of virtual boundaries in knowledge and organization. In: *North American Symposium on Knowledge Organization, 4*, Milwaukee, 2013, *Proceedings of the North American Symposium on Knowledge Organization*, p. 127-135. Available: <http://iskocus.org/nasko2013-proceedings.php>.

JAENECKE, P. To what end knowledge organization? *Knowledge organization*, Wuzburg, v. 21, p. 3-11, 1994.

KASTEN, J. Knowledge strategy and its influence on knowledge organization. In: *North American Symposium on Knowledge Organization, 1, 2007, Toronto, Proceedings of the North American Symposium on Knowledge Organization*, p. 44-54. Available: <http://iskocus.org/nasko2007-proceedings.php>.

KEMP, R.B. Classifying marginalized people, focusing on natural disaster survivors. In *North American Symposium on Knowledge Organization, 1, 2007, Toronto, Proceedings of the North American Symposium on Knowledge Organization*, p. 55-62. Available: <http://iskocus.org/nasko2007-proceedings.php>. 15

KIEL, E. Knowledge organization needs epistemological openness. *Knowledge organization*, Wurzburg, v. 21, p. 148-52, 1994.

KIPP, M.E.I. Tagging for health information organisation and retrieval. In: *North American Symposium on Knowledge Organization, 1, 2007, Toronto, Proceedings of the North American Symposium on Knowledge Organization*, p. 63-74. Available: <http://iskocus.org/nasko2007-proceedings.php>.

LA BARRE, K. The Heritage of Facet Analysis in North America: Past Lessons as Pathways for Contemporary Exploration. In: *North American Symposium on Knowledge Organization, 2, Syracuse, 2009, Proceedings of the North American Symposium on Knowledge Organization*, p.17-26. Available: <http://iskocus.org/nasko2009-proceedings.php>.

LOEHRLEIN, A. An examination of interdisciplinary theory between cognitive categorization and knowledge organization. In: *North American Symposium on Knowledge Organization, 3, Toronto, 2011, Proceedings of the North American Symposium on Knowledge Organization*, p. 122-129. Available: <http://iskocus.org/nasko2011-proceedings.php>.

LÓPEZ-HUERTAS, M.J. Some current research questions in the field of knowledge organization. *Knowledge organization*, Wurzburg, v. 35, p. 113-36, 2008.

MACCALL, S.L. Knowledge Organization under digital inversion: A theory for cooperative librarian organizing practices for online textual artifacts. . In: *North American Symposium on Knowledge Organization, 3, Toronto, 2011, Proceedings of the North American Symposium on Knowledge Organization*, p. 13-22. Available: <http://iskocus.org/nasko2011-proceedings.php>.

MARCHESE, C.; SMIRAGLIA, R.P. Boundary Objects: CWA, an HR Firm, and Emergent Vocabulary. In: *North American Symposium on Knowledge Organization, 4, Milwaukee, 2013, Proceedings of the North American Symposium on Knowledge Organization*. Available: <http://iskocus.org/nasko2013-proceedings.php>.

MARTÍNEZ-ÁVILA, D.; SMIRAGLIA, R. Revealing Perception: Discourse Analysis in a Phenomenological Framework. In *North American Symposium on Knowledge Organization, 4, Milwaukee, 2013, Proceedings of the North American Symposium on Knowledge Organization*. Available: <http://iskocus.org/nasko2013-proceedings.php>.

MCTAVISH, J.; FORTIER, A. A domain-analytic perspective on exual health in LCSH and RVM. In: *North American Symposium on Knowledge Organization, 3, Toronto, 2011, Proceedings of the North American Symposium on Knowledge Organization*, p. 83-93. Available: <http://iskocus.org/nasko2011-proceedings.php>.

MILANI, S. O.; GUIMARÃES, J. A.C. Biases in knowledge representation: an analysis of the feminine domain in Brazilian indexing languages. In: *North American Symposium on Knowledge Organization*, 3, Toronto, 2011, *Proceedings of the North American Symposium on Knowledge Organization*, p. 94-104. Available: <http://iskocus.org/nasko2011-proceedings.php>.

MOULAISON, H.L.; DYKAS, F.; BUDD, J.M. The Author and the Person: A Foucauldian Reflection on the Author in Knowledge Organization Systems. In: *North American Symposium on Knowledge Organization*, 4, Milwaukee, 2013, *Proceedings of the North American Symposium on Knowledge Organization*. Available: <http://iskocus.org/nasko2013-proceedings.php>.

MOYA-ANEGÓN, F.; HERRERO-SOLANA, V. Análisis de dominio de la revista mexicana de investigación bibliotecológica. *Información, cultura y sociedad*, [S.l.]: n. 5, p. 10-28, 2001.

OLSON, H. A.; HOWARTH, L.C. (2013). A Space of Transition: Rethinking Surrogates. In: *North American Symposium on Knowledge Organization*, 4, Milwaukee, 2013, *Proceedings of the North American Symposium on Knowledge Organization*, p. 161-166. Available: <http://iskocus.org/nasko2013-proceedings.php>.

PAJEK. *Introductory social network analysis with Pajek*. School of Information University of Michigan. Available at <http://ocw.mit.edu/courses/economics/14-15j-networks-fall-2009/assignments/MIT14_15JF09_pajek.pdf> Access at 16 jul. 2014.

PATTUELLI, M.C. Mapping subjectivity: performing people-centered vocabulary alignment. In: *North American Symposium on Knowledge Organization*, 3, Toronto, 2011, *Proceedings of the North American Symposium on Knowledge Organization*, p. 13-22. Available: <http://iskocus.org/nasko2011-proceedings.php>.

PIMENTEL, D.M. Exploring classification as conversation In:). In: *North American Symposium on Knowledge Organization*, 1, 2007, Toronto, *Proceedings of the North American Symposium on Knowledge Organization*, p. 1-8. Available: <http://iskocus.org/nasko2007-proceedings.php>.

PIMENTEL, D.M. Towards a typology of classificatory change. . In: *North American Symposium on Knowledge Organization*, 3, Toronto, 2011, *Proceedings of the North American Symposium on Knowledge Organization*, p. 13-22. Available: <http://iskocus.org/nasko2011-proceedings.php>.

PIMENTEL, D.M. The KO roots of Taylor's Value-Added Model. In: *North American Symposium on Knowledge Organization*, 2, Syracuse, 2009, *Proceedings of the North American Symposium on Knowledge Organization*, p. 58-67. Available: <http://iskocus.org/nasko2009-proceedings.php>.

SALES, R. GUIMARÃES, J. A. C. Contributions of Julius Otto Kaiser to the theoretical framework of Knowledge Organization: basis for the analytical-synthetic method. In: *North American Symposium on Knowledge Organization*, 4, Milwaukee, 2013, *Proceedings of the North American Symposium on Knowledge Organization*, p. 167-175. Available: <http://iskocus.org/nasko2013-proceedings.php>.

SCHATZ, J.; STENNES-SPIDAHL, N.; MILLS, S.; LOEHRLEIN, A. J. Bibliographic Induction: How KO Systems Optimize Browsing by Supporting Library Users' Prior Knowledge. In: *North American Symposium on Knowledge Organization, 4*, Milwaukee, 2013, *Proceedings of the North American Symposium on Knowledge Organization*, p. 176-186. Available: <http://iskocus.org/nasko2013-proceedings.php>.

SMALL, H. On the shoulders of Robert Merton: towards a normative theory of citation.. *Scientometrics*, v.60, n.1, p.71-79, 2004.

| 67

SMIRAGLIA, R. P. Performance works: Continuing to comprehend instantiation. In: *North American Symposium on Knowledge Organization, 1*, 2007, Toronto, *Proceedings of the North American Symposium on Knowledge Organization*, p. 75-86. Available: <http://iskocus.org/nasko2007-proceedings.php>.

SMIRAGLIA, R.P. Modulation and Specialization in North American Knowledge Organization: Visualizing Pioneers In: *North American Symposium on Knowledge Organization, 2*, Syracuse, 2009, *Proceedings of the North American Symposium on Knowledge Organization*, p. 35-46. Available: <http://iskocus.org/nasko2009-proceedings.php>.

SMIRAGLIA, R.P. Epistemology of domain analysis. In: SMIRAGLIA, R. P.; LEE, H-L, (Eds.). *Cultural frames of knowledge*. Würzburg: Ergon-Verlag, pp. 111-24.

SMIRAGLIA, R.P. Is FRBR A Domain? Domain Analysis Applied to the Literature of The FRBR Family of Conceptual Models. In: *North American Symposium on Knowledge Organization, 4*, Milwaukee, 2013, *Proceedings of the North American Symposium on Knowledge Organization*. Available: <http://iskocus.org/nasko2013-proceedings.php>.

SMIRAGLIA, R.P. Domain coherence within Knowledge Organization: People, Interacting Theoretically, Across Geopolitical and Cultural Boundaries. In: *Proceedings of the 39th Annual CAIS/ACSI Conference*, University of New Brunswick.

SOERGEL, D. *Dokumentation und Organisation des Wissens: Versuch einer methodischen und theoretischen Grundlegung and Beipiel der Sozialwissenschaften* [Documentation and Organisation of Knowledge : Towards a Methodological and Theoretical Foundation for the Social Sciences] Berlin, DEU: Dunker & Humbolt; 1971. 380 p.

SZOSTAK, R. Classification for Diversity. In: *North American Symposium on Knowledge Organization, 4*, Milwaukee, 2013, *Proceedings of the North American Symposium on Knowledge Organization*. Available: <http://iskocus.org/nasko2013-proceedings.php>.

TENNIS, J.T. Epistemology, theory, and methodology in knowledge organization. *Knowledge organization*, Würzburg, v. 35, p. 102-12.

TENNIS, J.T. The economic and aesthetic axis of information organization frameworks (extended abstract). In: *North American Symposium on Knowledge Organization, 1*, 2007, Toronto, *Proceedings of the North American Symposium on Knowledge Organization*, p. 112-113. Available: <http://iskocus.org/nasko2007-proceedings.php>.

TENNIS, J.T. Two Axes of Domain Analysis. *Knowledge Organization*, Würzburg, v. 30, n.3/4, p.191-195, 2003.

THORNTON, K. Contentious categories. In: *North American Symposium on Knowledge Organization*, 3, Toronto, 2011, *Proceedings of the North American Symposium on Knowledge Organization*, p. 217-226. Available: <http://iskocus.org/nasko2011-proceedings.php>.

TOGNOLI, N. B.; GUIMARÃES, J. A. C.; TENNIS, J. T. Diplomatics as a methodological perspective for archival knowledge organization. In: *North American Symposium on Knowledge Organization*, 4, Milwaukee, 2013, *Proceedings of the North American Symposium on Knowledge Organization*, p. 217-225. Available: <http://iskocus.org/nasko2013-proceedings.php>. 8

URBIZAGASTEGUI ALVARADO, R. Elitismo na literatura sobre a produtividade dos autores. *Ciência da Informação*, Rio de Janeiro: v.38, n.2, Maio/Ago. 2009.

VAN DEN HEUVEL, C. Multidimensional classifications: Past and future conceptualizations and visualizations. In: *North American Symposium on Knowledge Organization*, 3, Toronto, 2011, *Proceedings of the North American Symposium on Knowledge Organization*, p. 105-121.

VANZ, S.A.S.; CAREGNATO, S.E. Estudos de citação: uma ferramenta para entender a comunicação científica. *Em questão: revista da Faculdade de Biblioteconomia e Comunicação da UFRGS*, Porto Alegre, v. 9, n. 2, p. 295-307, jul./dez. 2003.

ZHANG, L.; KIPP, M.; LEE, H.-L. Genre Tag ... Tag Genre. In: *North American Symposium on Knowledge Organization*, 4, Milwaukee, 2013, *Proceedings of the North American Symposium on Knowledge Organization*.

ZINS, C. Knowledge organization: an epistemological perspective. *Knowledge organization*, Wurzburg, v. 31, p. 49-54, 2004.

The relation between knowledge organization and information science in the Brazilian scientific community: an investigation within ISKO-Brazil



39

Rodrigo de Sales

Federal Fluminense University
rodrigosaes@id.uff.br

1 Introduction

Organizing information to better retrieve it seems to be the motto of Information Science since its inception (BORKO 1968; SARACEVIC, 1996, SMIT & BARRETO, 2002). However, when the subject information organization is combined with the subject knowledge organization, some different viewpoints emerge and disturb the studies of the area. To Guimarães (2008), information organization, as a field of study that integrates Information Science, is one of the investigative spaces of this science, serving as mediaor in that it provides communication between the contexts of information production and use. For Bräscher and Café (2010), information organization relates to the activities and procedures concerning the material organization of information, the organization of informational items in the information units.

In this sense, the information organization is fundamentally articulated with the practices developed by Library and Information Science. The same does not seem to occur with equal quietness when it comes to knowledge organization. According Bräscher and Café (2010), knowledge organization regards the cognitive organization and systematization of knowledge, the organization of concepts as well as the construction of knowledge organization systems.

Although the authors place knowledge organization within Information Science, as does the organization of working groups of the National Association for Research and

Post-Graduation in Information Science (ANCIB), other concepts seem to exclude this belonging relationship where knowledge organization would not figure as something within Information Science. Perhaps the most explicit cases in which knowledge organization does not necessarily belong to Information Science are found in Dahlberg (1993, 1995, 2006 and 2014) and Hjørland (2003, 2008), who seek to assign knowledge organization with certain autonomy either as a discipline or a field of study or even as a new science.⁷⁰

This study aims to explore this detachment of knowledge organization (KO) in relation to Information Science (IS), i.e., the view that addresses KO and IS as two separate areas.

Thinking of knowledge organization developed by Brazilian researchers, which is investigated now is the understanding of how the Brazilian scientific community addresses the relationship between KO and IS.

Therefore, we used content analysis techniques, defined by Laurence Bardin, to analyze the papers published in the main forum for debate on knowledge organization in Brazil: Brazilian Chapter of ISKO.

2 Two perspectives at the international level: Dahlberg and Hjørland

The first decade of the twentieth century revealed the predominance of a perspective that defines KO as a "doing" of "operational" nature (GARCIA, OLIVEIRA, LIGHT, 2000; GREEN, 2002; GARCIA GUTIERREZ, 2002), whose research objects are mainly the concepts and conceptual frameworks (KENT, 2000; GREEN, 2002; OHLY, 2008 SMIRAGLIA, 2010), instrumentally formalized in Knowledge organization systems - KOS (KENT, 2000; GREEN, 2002; ZHEREBCHEVSKY, 2010; SOUZA; TUDHOPE and ALMEIDA, 2010). In this sense, we note KO linked to activities related to Information Science and Librarianship, more specifically to the thematic of information processing (FOSKETT, 1973).

However, concerning the nature (condition) of knowledge organization, Dahlberg (1993, 1995, 2006 and 2014) and Hjørland (2003, 2008), though with different perspectives, are authors who maintain, under the International Society for knowledge

organization (ISKO), a discourse of knowledge organization as an autonomous field of study. This view of KO as field of study can be found, to name a few, in Ohly (2012), Guimarães, Oliveira & Gracio (2012) and Barros & Moraes (2012).

Although distinct, the perspectives regarding KO presented herein are not necessarily mutually excluding and assist in the epistemological understanding of this scientific space, which although "young", is no longer in its embryonic state. ⁷¹

In the late twentieth century, the discourse of a new field was based on the statements of Dahlberg (1993, 1995), in which the author defined the scope, professional making, the institutional aspects, the classification of specialized literature and trends in KO. In the early twenty-first century, Dahlberg, who remains a central figure in the discourse that seeks to consolidate KO as an autonomous and independent area, tries to advance such perspective by assigning to KO the status of "scientific discipline" and/or "science."

In 2006, Dahlberg addresses knowledge organization as an autonomous and more comprehensive space than that traditionally linked to libraries and information units. Therefore, the author prefers the term "scientific discipline" and even suggests that finding a working institution formed by scientists, knowledge organizers and terminologists dedicated to the collection, the definition and systematization of concepts of all subject fields through a categorization structure formalized by a information classification coding. The author locates, within a universal system of science, KO as a subfield of Science of Science.

Dahlberg (2006) also outlines KO as a possible new science. The author makes use of distinctions made by Alwin Diemer to risk addressing KO as a new science, with its own objects, methods, actions and content.

Years later, Dahlberg (2014) reaffirms these ideas and emphasizes the need to remove the knowledge organization (KO) from Library and Documentation to accommodate it within the Science of Science, and also to other domains which deal with taxonomic inputs such as Zoology, Botany and Microbiology. In 2014, the author reclaims the formation of an institute or academy for KO with scientists and experts engaged in a joint work that could give fruitful results to the development of KO and science as a whole. In both articles, Dahlberg (2006 and 2014) even mentions Information Science (IS),

making her view of autonomous and independent KO clear, completely detached from IS. For the author, KO definitely should not be connected only to sciences related to information, but rather to science related to all fields of knowledge - the science of science.

Also in the early twenty-first century, Hjørland (2003, 2008) brought to KO the debate on the distinction between cognitive knowledge organization and the social knowledge organization. On the cognitive perspective, Hjørland (2003, 2008) stated it is knowledge organization addressed especially by Library and Information Science, whose focus fell on studies relating to the processes and instrumental constructions of KO, such as document description, classification and indexing.

On the social organization of knowledge Hjørland (2003, 2008) claims it is basically the organizations and categorizations of professions and disciplines, i.e., the social division of mental labor, such as the organization of university curricula and also classifications of subjects that divide and relate the different disciplines as well as the official tables formalizing the occupations of a country.

Concerning the interdisciplinary aspect of KO, Dahlberg (1993) proposed the KO's approximation with social areas (education, politics, industry and sociology), stating it is an interdisciplinary field that refers at least, to philosophy and terminology. Hjørland (2008) points out KO is not an exclusive subject of Library and Information Science, but also of areas such as Computing, Linguistics and Natural Language Processing, Theory of Knowledge, Theory of Social Organization and Metaphysics/Ontology.

It can be said we are far from a consensus on the effective epistemological connection between the IS and KO. Either we have a tradition, belonging to Library and Information Science, which prefers to address KO as fundamentally discussed under the theme of information processing, which is characterized as an investigative space belonging to IS, supported, for example, by studies by Foskett (1973); or we find a perspective that refuses this belonging relationship, claiming that IS would be, along with Librarianship, central discipline that would support KO issues (notably, the cognitive knowledge organization) - Hjørland. We also have a view that definitely decouples KO to IS - Dahlberg.

3 KO research in the Brazilian scientific context

In Brazil, KO studies initially found place in the research developed by the Brazilian Institute of Information in Science and Technology (IBICT), in the 1970s. However, it was in ANCIB's Working Group WG2 - Knowledge Organization and Representation, that Brazilian researchers were able to leverage and strengthen the development of KO in the country. ANCIB's working groups represent large specialized themes studied in the field of Information Science. In this context, KO figured as a theme, or subfield of Information Science.

In the two editions of ISKO's Brazilian Chapter (2011 and 2013), studies by Brazilian and foreign researchers dedicated to the theoretical and practical development of KO were presented. The Brazilian Chapter of ISKO has been consolidated as the main forum for discussions on KO in the country.

Evidently, KO research in Brazil is also based on the international literature of Knowledge Organization which, as shown, shows at least three different perspectives concerning their relationship with IS:

Perspective 1: evidences the intention of independence as a scientific discipline, positioning itself as a subfield of a Science of Science. When Dahlberg seeks to define the object, the activities, methods, researchers' goals, the classification of the discussed subjects, the creation of an international society, and when she makes use of terms such as "knowledge organizer" to refer to KO researchers, "field of study" or "scientific discipline", referring to the intended space for KO, she provides evidence of intention to address knowledge organization as field or independent discipline to IS;

Perspective 2: concerns IS matters (especially those related to information retrieval and information technology) to solve only part of KO, more specifically, the part on cognitive knowledge organization, such as those issues concerning units of information. This perspective places IS as part of KO, which, in turn, is concerned with broader issues. Hjørland, main supporter of this perspective, seems to want to detach KO as belonging to a theme or subfield to IS;

Perspective 3: traditionally places KO as part of IS and does not appear to seek independence, but to contribute to the central space of IS. This view finds support in the tradition that puts KO in direct dialogue with the thematic information organization,

developed in and for the information environment. The structural organization of ANCIB's working groups corroborates this view, since KO figures as one of the working groups inserted into IS.

In this research, we used these three perspectives as a basis to guide research aimed to understand how ISKO-Brazil researchers have been relating KO and IS. | 74

4 Application of the analysis and results presentation

The scope for forming the corpus of analysis was defined by all the texts published by Brazilian researchers in the Annals of ISKO-Brazil to date, 2011 and 2013. Therefore, in a universe of 86 publications, considering that 19 of them are of foreign authorship, the corpus of analysis was defined in 67 Brazilian authored publications. It should be noted that although all 67 publications have been analyzed, not all of them mention aspects concerning the nature or the relationship of KO with IS, and therefore are not shown or counted in the results. The analysis included only the texts that at some point refer to the nature and/or relationship of KO with IS, which totaled 20 texts. That is, out of the 67 texts that were analyzed, only 20 (29.8%) reported some information about the nature of KO and/or about the relationship between KO and IS.

As the classification measure of collected data, Bardin (2003) suggests the categorization process, which consists of grouping classes of converging elements in its characteristics. The defined categories were: KO Nature and Relationship between KO and IS.

To organize the analysis of information extracted from texts, each category was driven by inference variables, which allowed to deeply investigate the ideas presented by the authors of the texts.

Following the objective of this paper, it was necessary, firstly, to observe how the authors define the nature of KO. Thus, inferences variables for the category KO Nature were: 1 - scientific territory; 2 - action; 3 - subject and; 4 - science.

That is, how many and which texts consider that KO is a *scientific territory*, such as for example, a field, a discipline, an investigative space, a domain or area of study. The combination of all these terms in the same variable does not mean to consider them as

with the same concepts, but merely a technique of grouping concepts that have similar characteristics. In this first case, respecting the terms used by the authors, who either named KO as a field of study, either considered it a discipline or even domain, we tried to preserve what is common to the adopted terms, i.e., the territorial issue - KO as a scientific territory (whether physical or abstract). The same applies to the definition of other variables: *action* - KO seen as a process, a set of practices, or even as an activity; *subject* - for those who consider it as a theme, or a line of research, or even as an object of study. The last variable in this category seems to be the most obvious, *science*, for those who consider KO as a new science.

The next step was to define the variables that allowed better organization and understanding information regarding the second category, *Relationship between KO and IS*:

perspective 1 - KO as a scientific territory, **autonomous and unrelated to IS**;

perspective 2 - KO as a scientific territory, **autonomous but related to IS**;

perspective 3 - KO as a **theme within IS**, i.e. **KO as a subfield to IS**.

On the first category (Figure 1), it is clear the fact that most of the texts, 50% (10 texts), consider that KO is a scientific territory, be it a field, a discipline, a domain or area of study. This shows that in the Brazilian reality, KO has achieved its own investigative space boundaries, overflowing the concept of a procedural dimension that considers it an action, 15% (3 texts), and a thematic dimension, which considers it as a subject, 15% (3 texts). It should be noted that the concept led by Dahlberg in the international arena, claiming KO to be a new science, does not seem to significantly echo in the context of ISKO-Brazil, for such concept was depicted in only one of the analyzed texts (5%).

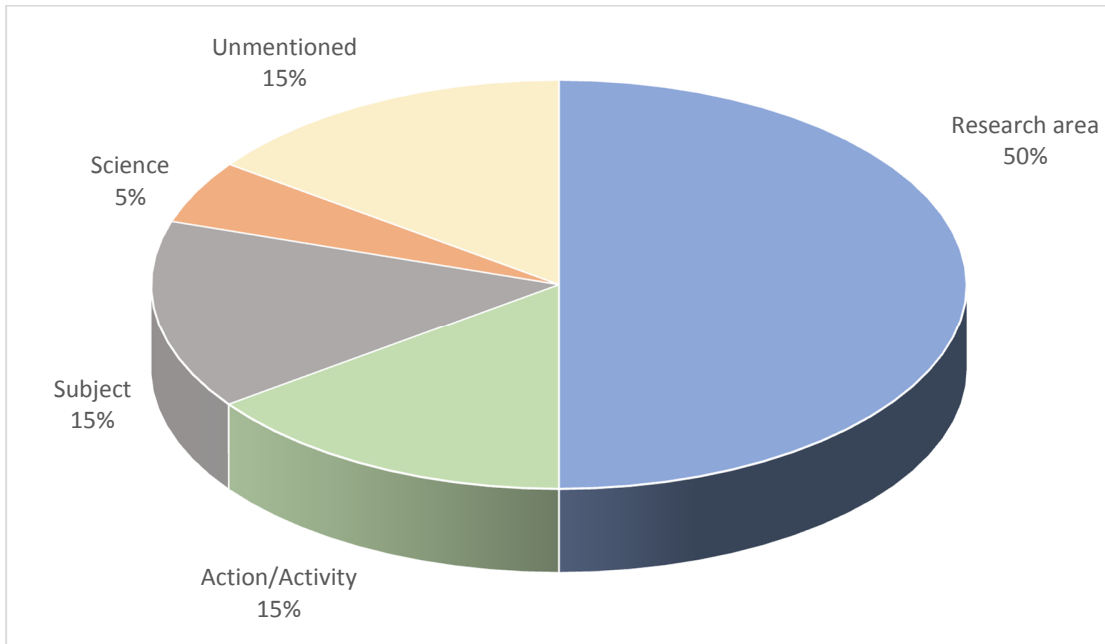


Figure 1: KO Nature (source: by the author)

Regarding the relationship between KO and IS (Figure 2), the main observation in this study (contemplated in category 2), is that most of the analyzed texts, 35% (7 texts), converge with perspective 2, understanding that KO is an autonomous investigative space, but strongly related with IS. Following this perspective, 4 texts (20%), believe that KO is actually a specialized topic or a subarea of IS, meeting perspective 3. Only one of the texts, 5%, converges with perspective 1, which considers KO an autonomous territory, unrelated to IS, or, as Dahlberg prefers it, that KO is a new science.

It is worth noting the percentage of the texts that did not mention KO nature (Figure 1) and the relationship between KO and IS (Figure 2). While 15% of the analyzed texts did not mention information about KO nature, 40% of them have not mentioned the relationship between KO and IS. This data shows how studies of this nature are increasingly needed for a better understanding of epistemological KO.

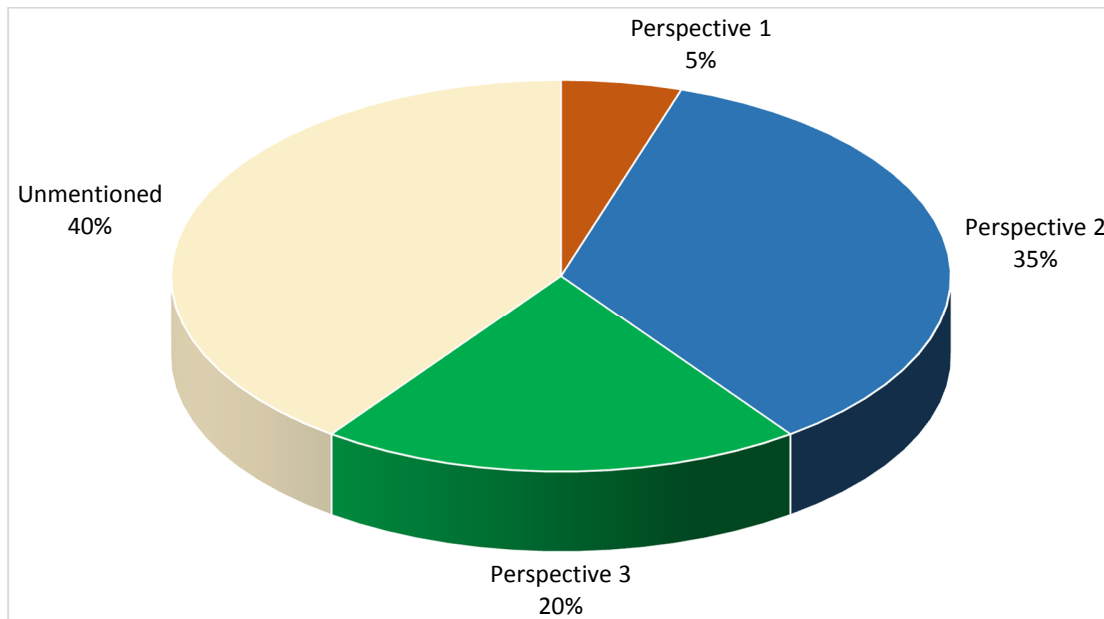


Figure 2 - Relation between KO and IS (Source: by the author)

Thus, the scenario that unfolds under ISKO-Brazil is that most Brazilian researchers understand that KO has achieved a status of scientific autonomous territory that is in constant dialogue with Information Science. In other words, KO, according to this scientific community, is no longer characterized only as a subject belonging to Information Science. The role of KO as CI subfield, as discussed by Brazilian research at the turn of last century to the twenty-first century, regarding its scientific status, is now divided with the idea that KO consists of a new territory overflowing the prospect initially developed in the country, and that it prints its own epistemological contours and is revealed as an emerging space.

5 Conclusion

The fact that IS is, in the turn of the twentieth century to the twenty-first, the area responsible for addressing and developing research relating to KO in Brazil did not prevent the latter to overflow the limits of that and gradually create their own boundaries. The institutionalization of KO through the Brazilian Chapter of ISKO, may have been both the cause and the consequence of this overflow in the Brazilian scientific community.

However, as this research showed, this new design, which seeks epistemological sedimentation and which already marks new space does not mean an absolute detachment of KO from IS. The bond that most Brazilians KO researchers have with IS, whether institutional, whether epistemological reflects a majority thought that prefers to understand both areas as spaces for essential dialogue and intersection, which denies the old belonging relationship.¹⁷⁸

References

The references was made following the ABNT rules.

BARDIN, L. *L'analyse du contenu*. 7ème. Paris: PUF, 2003. 296 p. (Le Psychologue, 69).

BARROS, T. H. B.; MORAES, J. B. E. Archival classification and knowledge organization: Theoretical possibilities for the archival field. In *Categories, contexts and relations in knowledge organization: Proceedings of the Twelfth International ISKO Conference 6-9 August 2012 Mysore, India*. Advances in Knowledge Organization, 13. Würzburg: Ergon, 2012, p. 272-276.

BORKO, H. Information science: what is it? *American Documentation*, v.19, n.1, p.3-5, Jan. 1968.

BRASCHER, M; CAFÉ, L. Organização da informação ou organização do conhecimento? In: LARA, M. L. G.; SMIT, J. (Org.). *Temas de pesquisa em Ciência da Informação no Brasil*. São Paulo: Escola de Comunicação e Artes/USP, 2010. ISBN: 978-85-7205-081-4.

DAHLBERG, I. Current trends in Knowledge organization. In: Garcia Marco F. J. (Org.). *Organización del conocimiento em sistemas de información y documentación*. Zaragoza: Universidad de Zaragoza. 1995, p. 7-25.

DAHLBERG, I. Knowledge organization: a new science? *Knowl. Org.* v. 33, n. 1, 2006, p. 11-19.

DAHLBERG, I. Knowledge organization: its scope and possibilities. *Knowl. Org.* v. 20, n. 4, 1993, p. 211-222.

DAHLBERG, I. What is knowledge organization. *Knowl. Org.* v. 40, n. 1, 2014, p. 85-91.
FOSKETT, A.C. *A abordagem temática da informação*. Tradução de Antônio Agenor Briquet de Lemos. São Paulo: Polígono; Brasília: Ed.UnB, 1973.

GARCIA, S. M. M.; OLIVEIRA; LUZ, G. M. S. Knowledge organization for query elaboration and support for technical response by the internet In: *Dynamism and stability in knowledge organization: Proceedings of the Sixth International ISKO Conference* Würzburg: Ergon. 2000. p.189.

GÁRCIA GUTIÉRREZ, A. L. Knowledge organization from a culture of the border: towards a transcultural ethics of mediation. In: *Challenges in knowledge representation and organization for the 21st century: integration of knowledge across boundaries: Proceedings of the Seventh International ISKO Conference*. Würzburg: Ergon, 2002, p.518.

GREEN, R. Conceptual universals in knowledge organization and representation In: Challenges in knowledge representation and organization for the 21st century: Integration of knowledge across boundaries: *Proceedings of the Seventh International ISKO Conference*. Würzburg: Ergon, 2002, p.15.

KENT, R. E. The information flow foundation for conceptual knowledge organization In: *Dynamism and stability in knowledge organization: Proceedings of the Sixth International ISKO Conference* 79 Würzburg: Ergon. 2002, p. 111.

GUIMARAES, J. A. C. A dimensão teórica do tratamento temático da informação e suas interlocuções com o universo científico da International Society for Knowledge Organization (ISKO). *Revista Ibero-Americana de Ciência da Informação*, v. 1, p. 77-99, 2008.

GUIMARÃES, J. A. C.; OLIVEIRA, E. T.; GRACIO, M. C. C. Theoretical referents in knowledge organization: A domain analysis of the knowledge organization journal. In *Categories, contexts and relations in knowledge organization: Proceedings of the Twelfth International ISKO Conference 6-9 August 2012 Mysore, India*. Advances in Knowledge Organization, 13. Würzburg: Ergon. 2012, p. 31-38.

HJORLAND, B. Fundamentals of knowledge organization. *Knowl. Org.* v. 30, n. 2, p. 87-111, 2003.

HJORLAND, B. What is knowledge organization (KO)? *Knowl. Org.* v. 35, n. 3/2, p. 86-111, 2008.

OHLY , H. P. Knowledge organization pro retrospective In: *Culture and identity in knowledge organization: Proceedings of the Tenth International ISKO Conference*. Würzburg: Ergon, 2008, p. 210.

OHLY , H. P. Mission, programs, and challenges of knowledge organization. In: *Categories, contexts and relations in knowledge organization: Proceedings of the Twelfth International ISKO Conference*. Mysore: Ergon, 2012, p. 15-23.

SARACEVIC, T. Ciência da Informação: origem, evolução e relações. *Perspectivas em Ciência da Informação*, Belo Horizonte, v. 1, n. 1, jan./jun. 1996, p. 42-62.

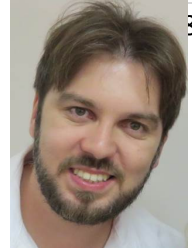
SMIRAGLIA, R. P. Perception, knowledge organization and noetic affective social tagging. In: *Paradigms and conceptual systems in knowledge organization*: Würzburg: Ergon, 2010, p.64.

SMIT, J. W.; BARRETO, A. A. Ciência da Informação: base conceitual para a formação profissional. In: VALENTIN, M. L. P. (Org.). *Formação do profissional da informação*. São Paulo: Polis, 2002. Cap. 1, p. 9-24.

SOUZA, R. R.; TUDHOPE, D.; ALMEIDA, M. B. The KOS spectra: a tentative faceted typology of knowledge organization systems. In: *Paradigms and conceptual systems in knowledge organization: Proceedings of the Eleventh International ISKO Conference*. Würzburg: Ergon, 2010, p.122.

ZHEREBCEVSKY. Formalism in knowledge organization. In: *Paradigms and conceptual systems in knowledge organization: Proceedings of the Eleventh International ISKO Conference*. Würzburg: Ergon, 2010, p. 98.

Poole, index and fractures: indexing and serial publications in the United States of the nineteenth century



Gustavo Silva Saldanha

Brazilian Institute for Information in Science and Technology
Universidade Federal do Estado do Rio de Janeiro
saldanhaquim@gmail.com



Naira Christofolletti Silveira

Federal University of the State of Rio de Janeiro
naira.silveira@unirio.br

Introduction to Fracture: from and to a bibliographic rationality

La Bibliographie des Périodiques et des Journaux est une oeuvre immense, elle ne peut être réalisée que par coopération et publiée sous forme de recueils divers. Elle est appelée à être concentrée dans le *Répertoire Bibliographique Universel*. (OTLET, 1934)

Between the entries *imprimeur* and *insectes* of the *Dictionnaire raisonné de bibliologie*, by Gabriel Peignot (1802), the absence of a term in the unfolding of alphabetical order suggests a great epistemic-historical question for informational studies: the development of indexing practices (in its inaugural sense of art of creating indexes) does not seem to occupy a central place in the bibliologic reflective thought that was born

there. Considering the exhaustive Peignot's review, we recognize the existence of the discussion on indexes, but not with the emphasis of a concept that deserves the uniqueness of an entry. Nor *index*, much less, of course, the concept of *indexation*, are highlighted in Peignot's compilation of the early 1800s. In fact, the speech of the French librarian directs to a prospective point of view - a science to be born - but relies in the eighteenth century as a vast source. The following century, "open" by the *Dictionnaire*, in turn, will see the rapid development, not only of indexing practices, but also the reflection on such concept with the dilemma of multiplication of textual artifacts and specialized languages.

The nineteenth century can be represented, among other characteristics, by the race for the creation of possibilities for a scientific language - one of the epistemological limits of Comtean positivism - which would allow the identification and construction of new empirical knowledge. This trajectory is accompanied by the development of theories, approaches and experiments focused on the construction of record forms and representation of record language, translated as knowledge records (objective index for "naturalization" potentiation of relations across knowledge, language and materiality along Modernity).

What we address as "fracture" in this document originally parts from two notions related to our purpose. The idea refers to senses of disruption, break, rupture, breach. It illustrates, in this reflection, this unfolding of specializations with a fragmentation of philosophy or general known, donor of forms and contents to others. In bibliographic vocabulary, "fracture" refers to another predominant meaning for this analysis: it is the Gothic script, characterized by the "contrast between thick and thin traces," which, having originated in the eleventh century, greatly influenced the twelfth and fourteenth, being in use even in the twentieth century, in Germany (FARIA, PERICÃO, 2008, p. 303).

The metaphor of the technical meaning of "fracture" is complementary and more objective to our discourse: we have a technical transformation in the bibliographic rationality exactly in the historical process that goes from the Cultural Renaissance to Artistic Renaissance, founding a world view in the first five centuries of the second millennium of our Western dating devoted to empiricism, to inductive landmarks and a "document mode" of making science. Along the various scientific specializations, the

demand for a specific subject, for a theory of knowledge, of metascientific character is established, or, in the visions of Peignot (1802) and Otlet (1934), we are facing another result of "fracture", the emergence (or the demand for construction) of Bibliology. Supplementary and not least important, another sense, coming from the context of a theory of bibliographic knowledge, is recognized: the fragmentations of the Book (as Otlet's view indicated, in capital letters) in mass production and, in these, the avalanche of countless chains of new discourses, knowledge, specialized languages.

To solve the problem of contiguous multiplication of languages and new instruments, namely, the new "discourse communities" that interact in certain "domains" (HJØRLAND, ALBRETSCHEN, 2005) in the course of the nineteenth century, the theoretical and practical interventions of craftsmen dedicated to the organization of knowledge is presented. It is necessary, given the myriad of languages, a range of instruments to meta-represent symbolic artifacts built by each community scattered in the landscape and closed in their lexicons, their instruments, their methods, such as indexes, bibliographies and catalogs.

Related to the portrayed spatiotemporal context, we recognize the presence of William Frederick Poole, as a central figure for the consolidation of an American thought in thematic retrieval of artifacts, especially in the context of journals with worldwide influence. Given the dynamics of mass production and the profusion of content that may come from it, we notice a systematic experience, the result of a dialogue between theory and practice, for the thematic representation of memory instruments and communication of distinct discursive communities that multiplied on the "knowledge fractures." Within a historical epistemology, interested in the recognition and discussion of concepts, methods, approaches and experiments in the discursive space-time of informational studies, the objective is to discuss, from a theoretical approach, the development of information retrieval tools in the context of mass production in the nineteenth century, with the Poole as an analytical focus of this context.

Treated Fractures: from index construction to index

The only comprehensive general index to periodical literature is that prepared by Mr. Poole [...] (ROBINSON, 1876, p. 665)

To discuss the bibliography of journals and newspapers, Paul Otlet (1934) highlighted the challenges inherent to the practice of descriptive and thematic representation of artifacts, that is, their complexity and demand not only for theories and methods for proper appropriation, but also national and international policies for their production and use. Issues such as standardization for abbreviations of journal titles, editors' data, breaking records, indication of libraries that store and provide such documents represent some of these demands.⁸³

Among the classical works devoted to this concern, Otlet (1934) emphasized the *Schedularium of Concilium Bibliographicum* and the *Catalogue of Scientific Papers*, in the nineteenth century, in addition to the *Catalogue International de la Littérature Scientifique*, this latter in the twentieth century, developed between 1900 and 1914. In this relation, the Belgian lawyer points out that "Le *Poole Index* a été la première oeuvre de cette nature" (OTLET, 1934, p. 296). It is curious to mention, since it is an American index, recognized as leading by an Old World theoretician concerning the understanding of knowledge dynamics re-characterized in disciplinary fractures. This concern is the clear indication of the political relationship between the development of an unborn indexing in the US territory, the intense production of recorded knowledge and a demand for access to fragmented knowledge.

The so-called Poole's index to Periodical Literature sought to identify, relate and represent the content of periodical documents of various types, aimed at their retrieval. It was, therefore, to some authors, "[...] first research instrument of this kind" (BATTLES, 2003, p. 142). According to Otlet (1934), the Poole's tool could register 590,000 articles from 470 different periodicals and was continued by the *Annual Literary Index* and by the *Annual Library Index*.

Because of its characteristics, it became a very important tool for the development of indexing and has been present to date. The fact is corroborated by Silva and Fujita (2004, p. 140), who report that the index published in 1882

[...] introduced the cumulative index of periodicals, using subject entries for periodical articles represented by keywords drawn from the titles of the

indexed articles [...] W.F. Poole was responsible for major periodical index of the nineteenth century.

Even after so many years, we observe that the reasons driving Poole to build his index remain unchanged, providing a tool to retrieve scattered information. According to what is stated by the following definitions:

| 84

INDEXES: Indexes provide subject, author, and/or title indexing to a particular set of periodicals and gives a full citation for each article. The citation includes the title of the periodical, date, volume, pages, as well as the author and title of the article. Some indexes include ABSTRACTS. (INDEXES, 2014, *online*)

index

Points to where information can be found.

List at the end of books, encyclopedias, etc. that indicates by author, title and/or subject the location of information within the book or encyclopedia. Tool that arranges (by author, title, or subject) citations to articles in a selected group of periodicals. (INDEX, 2014, *online*)

Although its origin is related to the series documents, specifically to periodicals, it is now common to find index on various types of documents or referring to more than one document. Periodicals can be regarded as continuous documents. They are often published for many years and each volume can encompass a series of articles on many subjects scattered in their numbers. Thus, it is defined "periodical publication: Publication in any medium, issued in successive physical units, with numerical and/or chronological designations and intended to be continued indefinitely" (ABNT, 2002, p. 2).

Due to the characteristics of periodicals, it is easy to understand why this type of document led to the creation of indexes. Some journals periodically publish their content in order to retrieve scattered and fragmented knowledge in their collection. The elaboration of indexes is related to the retrieval of dispersed information, so this instrument is often related to Documentation, although it has appeared in a Librarianship environment. For Lancaster (1993, p. 1), "The main purpose of the development of indexes and abstracts is construer document representations published in a way that lends itself to inclusion in some kind of database." At Poole's time, this database was a printed document, prepared with the best technology of that moment.

Indexes are composed of terms extracted from a document through the indexing process. Such terms are transformed into descriptors of content that will be used in information search and retrieval (LANCASTER, 1993).

The indexing process is related to subject representation, Subject Indexing of is an expression used in a more imprecise way; it refers to the representation of thematic contents of parts of complete bibliographic items, as is the case of the index at the end of a book. The process by which the thematic content of bibliographic items is represented in published databases - in print or computer-readable form - is almost invariably referred to as subject indexing, whether they are examining items globally or partially (LANCASTER, 2003, p. 15).⁸⁵

These recognized definitions make up the vast framework of indexing theory. This, together with its uses, experiments, instruments, is the cause and the result of socio-political transformations, which had an accelerated advance in the nineteenth century, following the multiplication of serial order artifacts. In this sense, the thought and the empirical unfolding of Poole's actions are a basic model for understanding the construction of this theory and its concepts.

Poole and its Index in Space-Time: indexing, policy and the construction of *Library Science* in the nineteenth century

"... and we recommend that a similar system of indexing be extended to the transactions and memoirs of learned societies." (CLAPP, 1954)
[Message Library Conference in the United States in 1853]

If the US Librarianship thought reverberates, in the world from the late nineteenth century, establishing itself as a major source of influence for the development of what it was renamed in the future of Library and Information Science, its unborn theoretical approaches of the nineteenth century have direct epistemological impact on this constitution. However, epistemology is never in a "pure" state, in a "social isolation" and "a-political" condition. Indexes have a foundational role (along with other approaches) in the American Librarianship thought.

In 1848, Poole, starting his career as a librarian at Yale University, developed the study on the journals of the university library, trying, instead of performing traditional lists,

to understand the issues involved in the various articles present in the serial artifacts. In the same year, Poole's manuscript draws attention and is published by George P. Putnam, "An Alphabetical Index to Subjects Treated in the Reviews, and Other Periodicals, to Which No Indexes Has Been Published: Prepared for the Library of the Brothers in Unity, Yale College." Poole's work is a systematic development of the studies by John Edmands,⁸⁶ former librarian at Yale, developed in 1947. (CLAPP, 1954)

After the first "impact", only in the "miraculous" year of 1876, the proposal is resumed. The central focus was on the cooperative development of index, with 51 libraries, focusing on description of 232 journals. The reception of the work, according to Clapp (1954), is clearly great: the British Museum is one of the first institutions to acquire the work and worry about its development. The context of the last three decades of the nineteenth century in the United States already marked a kind of "indexing industry," or a set of individuals, public and private institutions dedicated to the "market of metarepresentations of thematic background".

From the data receipt of each library about their journals, the librarian would redistribute the metarrepresentations for all cooperating ones.

O projeto, em poucas palavras, foi este: eu imprimiria e enviaria para as principais bibliotecas uma lista de periódicos que deveriam constar no índice. As bibliotecas verificariam a presença dos periódicos e retornariam com as listas particulares. Tendo recebido estas listas, eu faria uma distribuição equitativa dos trabalhos, tendo uma cota para mim, redistribuindo proporcionalmente os itens para as bibliotecas maiores e menores. Cada biblioteca elaboraria seu índice de acordo com um código de regas fornecidas e enviaria as referências para mim, que iria revisar, organizar e incluir. (POOLE, 1882, tradução nossa)

Cooperative work had repercussions in the United Kingdom, and the exchanges were also established across the Atlantic. Poole (1882) points out, in the theoretical and technical discussion of developing index, guidelines such as standardization, scope, quality of descriptors, abbreviations, dating. The indexer also draws attention to the general purpose of the index: recognizing the demand of "dispersed discursive communities," Poole (1882) makes it clear that the general purpose of the index is to reach students, journalists and professionals of the book world, and thus having a general look, while recognizing the specialized languages. In this sense, Poole (1882) points to indexing

experiences of scientific papers and legal and health areas. The "fractures of scientific knowledge" reverberate in the everyday world, and because of that the "new subjects" represented in the Index. While not the focus, scientific and professional journals were included when their thematic objectively approached the general elements of ongoing transformations. Poole's Index is, thus, direct influence on the immediate demands of specialized information retrieval. ¹⁸⁷

Some final words

The "public library" which we are to consider is established by state laws, is supported by local taxation and voluntary gifts, is managed as a public trust, and every citizen of the city or town which maintains it has an equal share in its privileges of reference and circulation. (POOLE, 1876)

In Poole's specific space-time, we realize the development of a close relationship between thematic retrieval of artifacts, democracy and epistemology of knowledge organization. In other words, we can recognize how, on one hand, under a specific theoretical scope of the future *American Library and Information Science*, i.e. indexing, establishes a chance to contribute to the American democratic ideal of universal access to knowledge in the 1800s, which co-results in the emergence of a specific field, a new epistemology (more an enhanced and emancipated domain) in the fracture of the general organization of the knowledge tree, "fragment which will be devoted to the fragmented".

References

The references was made following the ABNT rules.

CLAPP, Verner W. Indexing and abstracting services of serial literature. *Library trends*, v. 2, p. 509-21, apr. 1954.

BATTLES, M. *A conturbada história das bibliotecas*. São Paulo: Planeta, 2003.

FARIA, M. I. R. de; PERICÃO, M. da G. *Dicionário do livro: da escrita ao livro eletrônico*. São Paulo: Edusp, 2008.

HJORLAND, B.; ALBRECHTSEN, H. Toward a new horizon in information science: domain-analysis. *Journal of the American Society for Information Science*, v. 46, n. 6, p. 400-425, 1995.

INDEX. In: UNIVERSITY OF ILLINOIS. University Library. *Glossary of library terms*. Urbana, IL: University of Illinois, 2014. Available at: <<http://www.library.illinois.edu/learn/intro/glossary.html#i>> . Access at: 01 maio 2015.

INDEXES. BOSTON COLLEGE. UNIVERSITY LIBRARIES. *Library Glossary*. Boston: Boston College, 2014. Available at: <<http://www.bc.edu/libraries/help/howdoi/howto/glossary.html#i>> . Access at: 01 maio 2015.

| 88

LANCASTER, W. F. *Indexação e resumos: teoria e prática*. Brasília: Briquet de Lemos, 2003.

POOLE, W. F. *Poole's Index to Periodical Literature*. Chicago, 1882. Available at: <<http://c19index.chadwyck.com/infoCentre/introvol1.jsp>>. Access at: 25 jan. 2013.

OTLET, Paul. *Traité de documentatation: le livre sur le livre: théorie et pratique*. Bruxelles: Editiones Mundaneum, 1934.

PEIGNOT, G. *Dictionnaire raisonné de bibliologie*, tomo I. Paris: Chez Villier, 1802.

POOLE, W. F. *An Index to Periodical Literature*. New York: C.B. Norton, 1853.

POOLE, W. F. The organization and management of public libraries. In: EUA. Bureau of Education. *Public libraries in United States of America: their history, condition, and management*. Parte I. Whashington: Government Printing Office, 1876. p. 465-476.

ROBINSON, Otis H. On indexing periodical and miscellaneous literature. In: EUA. Bureau of Education. *Public libraries in United States of America: their history, condition, and management*. Parte I. Whashington: Government Printing Office, 1876. p. 663-672.

SILVA, M. dos R. da; FUJITA, M. S. L. A prática de indexação: análise evolutiva de tendências teóricas e metodológicas. *Transinformação*, Campinas, v. 16, n. 2, p. 133-61, 2004. Available at: <<http://periodicos.puc-campinas.edu.br/seer/index.php/transinfo/article/view/717>>. Access at: 01 maio 2015.

Processes of archival knowledge representation: historical and conceptual elements for classification and description



Natália Bolfarini Tognoli
São Paulo State University
nataliatognoli@marilia.unesp.br



Thiago Henrique Bragato Barros
Federal University of Pará
thiagobarros@ufpa.br

39

Introduction

Classification and description are nuclear activities in the process of organizing and representing archives. This knowledge can be understood as all knowledge generated from the activities carried out by individuals or legal entities that result in organic information.

In Archival Science, the principle of *Respect des Fonds* - or Principle of Provenance, since the second half of the nineteenth century, has been considered the most important in the processes of organization and representation. Established in 1841, due to the need of the French government to solve problems of uncontrolled accumulation of documents in the National Archives after the French Revolution, the principle states that the documents produced or accumulated by a person or institution should be grouped together and kept together, thus generating the so-called archive funds.

For Duchein (1983, p. 64) "the simplest definition of *respect des fonds* means to group, without mixing them with others, the archives (documents of every kind) created by or coming from an administration, establishment, person or corporate body."

In this sense, it is understood that the documents collected in funds and/or collections reflect the knowledge produced on a particular person or specific institution, involving three main actors in this context: (1) the creator (author), the person or entity responsible for the creation of documents; (2) the user, who will use the document for

evidence, administrative, legal, social, historical and/or scientific purposes, and (3) managers, archivists, documentalists, i.e., the actors somehow responsible for the organization and socialization of this information.

Therefore, it is believed that it is possible to understand the archival representation work as a form of knowledge organization, i.e., the archivist and Archival Science as an area with an interdisciplinary relationship with Knowledge Organization, as defined by Hjørland (2008, p.86) as "activities such as document description, indexing and classification performed in libraries, bibliographical databases, and archives [...]."⁹⁰

We present the historical and conceptual elements of the representation process, focusing on the activities of classification and description as fundamental for organization and access to knowledge generated from organic information. To do so, we follow the classification logic as the first activity in the representation process, preceding description.

The activities of description and classification are, for the archives, whichever are the activities of classification, cataloging and indexing for libraries, i.e., the organization of archival knowledge itself.

The aim of this historical and conceptual path is to approximate typical archival processes to the theory developed in the theoretical-conceptual space of Knowledge Organization.

Archival classification: key elements for knowledge organization

Classification in the archives reality is considered a borderline activity as it bridges the activities of management and the activities of access and document preservation.

The activities of classification, as well as the description ones, were the first activities to be in some way conceptualized in theory and archival practice, developed from the nineteenth century.

The basic principles of provenance and original order will establish from the end of the nineteenth century as fundamental for archives organization.

In the early twentieth century, classification advances with reconsiderations regarding the theme with Hillary Jenkinson's publication in 1922. For the author, classification is divided into two parts: the first part is concerned with the management of

the institution, its history and organization and the second divides the files into classes and their subdivisions (JENKINSON, 1922, p 81).

Classification, as a theory, develops to the fullest from the 1950-1960 decade with the publications of Theodore R. Schellenberg, Oliver W. Holmes, Ernest Posner, Petter Scott, among other archivists and researchers from the United States and Australia's⁹¹ National Archives.

While the discipline progresses deeply, from the 1950s, a break that will be perpetuated over the decades to the present takes place. This rupture concerns the separation of certain activities and concepts determined by the phase the documents are.

One of the institutions responsible for that is the American National Archives, centered on the figure of Theodore R. Schellenberg, who spread and popularized his methods, separating the classification as a typical activity of active archives from the arrangement of permanent archives, initiating a systematic rupture in the discipline.

It is noted that before that, there was no difference between these functions and the most commonly used term, arrangement, was used to assign things which, for the American Archival Science, after the 1950s, are different. Eastwood makes a fundamental consideration regarding this terminological and conceptual problem. The author claims that the choice for the term "arrange" to name the process is unhappy. It implies places things in an acceptable way, in a convenient order, such as arranging books in a bookshelf. The term classification is no longer satisfactory, as in this case it implies the arrangement or ordering of things in classes and it is better used in archival science for the process of organizing active documents (EASTWOOD, 2000, p. 93).

Classification by function is today, at the same time, the asset and the simulacrum of Archival Science, in the sense that for some authors is by that perspective that the activity is defined, such as, for example, Shepherd and Yeo (2003, our translation) "classification schemes are based on the analysis of the functions, processes and activities"

Classification is a managerial and planning activity, not only for archives, but also for administrations that generated the documents. Thus, the construction of the classification scheme should include the relations between documents, not only these,

but also the relationship between the documents and the people, thus the classification system would become a useful and facilitator tool of document description and appraisal.

Thus, classification serves as the basis for every archival organization process and obviously for archival knowledge representation.

Archival Description: historical and conceptual elements

The word description is derived from the Latin term *descriptio*, originated from the term *describere*, which means to transcribe, copy, narrate, define, distribute, put into classes, write about. Therefore, the term archival description literally means **to write about the archival material** and comprehends the representation, identification and organization of ideas.

This activity of describing archival documents has been present since ancient times, and their goals have changed over time.

In ancient times, for example, repertoires of documents were built on clay tablets, and served to avoid direct consultation to the documents, and keep the contents of archives in case of sudden shifting or in case of war or fire. In this case, the act of describing was linked to the preservation of records.

In the Middle Ages, inventories were compiled in order to provide evidence of document existence; or store the preserved material for custodians successors. Here, in addition to preserving the documents, the description had a legal objective (as it provided proof of documents' existence), and administrative, as it enabled the custodian of documents to maintain control over the funds and facilitate information retrieval for value primary purposes).

The need for physical and administrative control was keenly felt, not only for the purpose of effective retrieval, an operation which was conducted mainly on the basis of location lists and the physical arrangement of the records themselves, but also for guaranteeing that the archival repository sewed its function of "perpetual memory." Perpetual memory is a juridical concept according to which the documents preserved in an archives are authentic and permanent evidence of past actions. This idea governed every archival endeavour until the eighteenth century, and was the main reason for the preservation as well as for the description of archival material (DURANTI, 1993, p. 49).

Although the act of describing documents has existed since ancient times, the term "archival description" and its linkage to classification, while two integrated activities, date from the nineteenth century, due to the historiographical and scientific movement of the time, which come to understand the archive as testimony to studies fundamentally related to medieval documents.

During this period, another objective is added to the activity of describing: to provide access to documents. With the opening of the archives to citizens after the French Revolution, it is noted that more people besides the creator and some academic scholars have begun to use the archival material. Therefore, adaptation and redirection of the description activity was deemed necessary, as Duranti highlights (1993, p. 50)

[...] description has progressively lost the purposes of accounting for the holdings of the archival institution, facilitating physical retrieval of documents and providing access to information. These purposes are now fulfilled by the compilation of accession registers, location lists, indexes and all sorts of retrieval aids. The primary purpose of writing about the records has gradually become in Europe (and is beginning to become in North America) that of illuminating provenancial and contextual relationships.

Therefore, from the twentieth century, the description starts to worry less with the creation of simple search tools and subrogates of documents, and starts to focus on the document's relationship to its context, working in parallel with classification. It can be said, then, that to describe is a process that allows the archivist to represent - in the most denotative sense of the term - all knowledge that is kept in its record.

For Rodrigues (2003, p. 217), description is a key role regarding the representation of archival information and the possibility of access to them. This representation goes beyond the document, including its **functions** and **producing organs**. In this regard, the principle of *respect des fonds* carries again a central role in establishing the levels that will be described.

The understanding of the description and classification as joint activities enables the existence of interdisciplinary relations between Archival Science and Knowledge Organization, allowing Archival Science new methodological theoretical approaches.

Conclusion

Archival Science has a recent theoretical development in relation to its conceptual foundations and methodological development. We sought to present integrating aspects of archival theory, open to relations with knowledge organization aiming at an appropriation by both areas, since the archives become a space for applying methodologies of knowledge organization and Archival Science gains a new space for discussing its historical and conceptual precepts.

Both areas have to gain from relations of this nature and studies such as this one should run more frequently and in depth targeting an increase in methodological apparatus in this universe.

References

The references was made following the ABNT rules

DUCHEIN, M. Theoretical principles and practical problems of respect des fonds in Archival Science. *Archivaria*, v. 16 (summer), p. 64-82, 1983.

DURANTI, L. Origin and development of the concept of Archival Description. *Archivaria*, v. 35 (Spring), p. 47-54, 1993.

JENKINSON, H. *A manual of archive administration: including the problems of war archives and archive making*. Oxford: The Clarendon Press, 1922.

HJØRLAND, B. What is Knowledge Organization (KO)? *Knowledge Organization*, v. 35, n. 2, p. 86-102, 2008.

RODRIGUES, G. M. A representação da informação em arquivística: uma abordagem a partir da perspectiva da norma internacional de descrição arquivística. In: RODRIGUES, G.M.; LOPES, I.L.(Org.). *Organização e representação do conhecimento na perspectiva da ciência da informação*. Brasília: Thesaurus, 2003, v. 2, p. 210-230.

SABOURIN, P. Constructing a functional-bases records classification system: business activity structure classification system. *Archivaria*, n. 51, p. 137-154, 2001.

SHEPHERD, E.; YEO, G. *Managing records: a handbook of principles and practice*. London: Facet, 2003.

Ontologies and Simple Knowledge Organization Systems (SKOS): similarities and differences



Rogério Aparecido Sá Ramalho
Federal University of São Carlos
ramalho@ufscar.br

Introduction

One of the main challenges inherent to the area of Knowledge Organization (KO) is the consensus on what to represent and how to do it, therefore, different areas (Information Science, Computer Science, Linguistics, Cognitive Science, among others) have presented several models and representation instruments, making its distinction more complex along the time, both theoretically and practically.

In the context of Information Science, in order to understand the real importance of research in KO, it is necessary to consider its own meaning and scope from studies performed over the years which have resulted in a large number of standards and guidelines, as well as in theoretical developments and advances in information technology (Hjørland, 2003).

Over the recent years, research related to the development of ontologies has favored approximation between the instruments traditionally used in the field of information science and new semantic technologies. "[...] ontologies have been identified as a new possibility for representation of information resources in libraries" (Ramalho, 2010, p. 17, our translation). In this context, from August 2009, with the publication of the World Wide Web Consortium (W3C) recommendation for the Simple Knowledge Organization System (SKOS), this new representation instrument has attracted the attention of researchers, sometimes being compared to ontologies, which contributed to

cause terminological and conceptual conflicts about existing approaches regarding such instruments.

Based on these assumptions, this study presents a brief interdisciplinary discussion on the literature, aiming to contribute to a greater terminological clarification on existing similarities and differences between ontologies and SKOS, analyzing the potential of each of these instruments and their effects on the theoretical corpus of Information Science.⁹⁶

Ontologies as representation instrument

Within the scope of Information Science, the term ontology started to be used in the late 1990s, when information representation instruments attracted greater interest of researchers from various fields of knowledge, driven by the creation of the Web environment and development of new digital technologies (SOERGEL, 1999; VICKERY, 1997).

García-Marco (2007) points out that the increasing number of searches related to ontologies and metadata in Information Science constitutes the "tip of the iceberg" of a disciplinary reconfiguration process, as a result of integration of various sciences. Thus, it is noticed that over the past decade a greater commitment to the consolidation of definitions of ontology and analysis of the possibilities of use as an instrument of representation has taken place (SANTOS; CORREA, SILVEIRA, 2013).

Among the most referenced ontology definitions in the literature, we highlight the one presented by Gruber (1993, p. 1) "ontology is an explicit specification of a conceptualization." However, this definition refers to the concept of ontology in philosophical aspects, favoring the generation of terminological conflicts as highlighted by Guarino (1998, p. 5):

In the philosophical sense, we may refer to an ontology as a particular system of categories accounting for a certain vision of the world. As such, this system does not depend on a particular *language*: Aristotle's ontology is always the same, independently of the language used to describe it. On the other hand, in its most prevalent use in AI, an ontology refers to an *engineering artifact*, constituted by a specific *vocabulary* used to describe a certain reality, plus a set of explicit assumptions regarding the *intended meaning* of the vocabulary words.

Ontologies appear as a new computational approach to representation, description and organization of informational content, favoring the approach of classical representation models, traditionally used within the area of Information Science, with new semantic technologies developed over the last few years. As the focus to semantic representations, Zafalon (2013) explains that "it is the semantics that addresses the mental processes by which one produces, constitutes, understands and describes the representation of an information resource."

Over the last years, several methods have been proposed for developing ontologies. Among them, we highlight the ones by Uschold and King (1995), Uschold and Gruninguer (1996), Fernández, Gómez-Pérez and Juristo (1997), and the method 101, proposed by Noy and McGuinness (2001). Within the area of Information Science, as theoretical substrates that can be used for the development of ontologies, we can highlight the theory of concept and the theory of faceted classification used in the preparation of thesauri (CAMPOS, GOMES, 2006; GOMES, et al ., 1990) as well as ANSI/NISO Z39.19-2005, which provides guidelines for the development of controlled vocabularies.

In order to bring the concept of ontology to the assumptions of Information Science, especially regarding the processes of representation, processing, organization and retrieval of information resources, we refer to Ramalho (2010, p. 107) that defines ontology as

theory of faceted classification [...] a technological artifact that enables to formally represent the properties and relations of a given conceptual model, favoring the use of automatic inferences in organizational and retrieval processes of information resources.

Such definition characterizes ontologies within Information Science as an epistemological level instrument, conceived with the purpose of enabling the realization of automatic inferences, not being enough to represent existing concepts and relations, but to ensure computability of information through inferences.

In order to ensure such features, ontologies are usually implemented from the Web Ontology Language (OWL), language developed in the wake of the Semantic Web project, to ensure maximum expressiveness without losing computational completeness and

which could be inserted into Web pages. OWL can computationally represent classes, properties, individuals and data values, which are stored as Web documents.

In the first version, OWL had, as main purpose, to allow the representation of classes, individuals and relations. More recently, with the publication of OWL 2, new features have been incorporated, allowing to express additional restrictions on properties,⁹⁸ new characteristics of properties, incompatibility of properties, property chains and key properties (W3C Recommendation, 2012).

Simple Knowledge Organization System (SKOS)

SKOS is a data model that enables expressing Knowledge Organization Systems (KOS), such as thesauri, classification schemes and taxonomies in a computer format "readable" by machines. SKOS was developed from successive projects and working groups of the World Wide Web Consortium (W3C) and adopted in 2009 as a W3C Recommendation to support representation, use and interoperability between different types of vocabularies (ZOGHLAMI; KERHERVÉ; GERBE, 2011).

From the use of SKOS, concepts can be identified using URIs, labeled with textual sequences in one or more natural languages, embedded notes, documented with various types of notes linked to other concepts organized into informal hierarchies and association networks, aggregated into conceptual schemes, grouped into labeled and/or ordered collections or mapped into concepts from other schemes (W3C Candidate Recommendation, 2009).

For the formalization of data represented in SKOS, the standard Resource Description Framework (RDF) is used. Despite having a limited expressiveness, RDF allows the formal representation of statements about objects by providing mechanisms for describing specific resource relationships, limited to instance level and not allowing the use of logical connectives (RAMALHO; OUCHI, 2011)

As an RDF application, SKOS allows concepts to be represented on the Web, linking data and integrating other conceptual schemes. Each SKOS concept is defined as an RDF resource and is linked with other concepts and properties. Thus, it can be stated that the SKOS model is based on a vocabulary vision, where the primary objects are not labels, rather, they are concepts represented by labels which can be linked to other metadata standards and representation schemes.

Conclusions

It is observed that the construction of ontologies seeks the formalization of relations among concepts that can be computationally processed, so that heterogeneous data structures may be "understood" computationally, favoring logical "interpretations" from automatic inferences.

Thus, the development of ontologies requires greater involvement of communities of experts, so that the information expressed in an ontology may computationally represent a consensus of specific communities.

Regarding the SKOS data model, it is characterized as an innovative way to represent in computer format the instruments and representation schemes traditionally used in the area of Information Science, such as thesauri, classification systems, and taxonomies. However, due to its primary purpose, which is to establish a simple vocabulary that favors mapping existing concepts in different conceptual schemes, the computational representation of the SKOS model is limited.

Thus, it is observed that SKOS's purpose is quite different from the desired ontologies' expressiveness and such instruments should not be confused. However, despite its limitations, the SKOS model stands out because of the potential offered to information resources interoperability. It is expected that the conceptual mappings from the SKOS model be one of the main contributions to the future of Semantic Web, linking different conceptual schemes.

The ontologies still appear as the most significant representation instrument capable of engendering a new epistemological "constellation" in the field of Knowledge Organization, pointing to innovative perspectives and new possibilities in the professional make of Information Science, as the area demands greater methodological rigor in establishing its formal relations in order to favor automatic inferences.

References

The references was made following the ABNT rules.

FÉRNANDEZ, M.; GÓMEZ-PÉREZ, A.; JURISTO, N. *Methontology: From Ontological Art Towards Ontological Engineering*. Palo Alto: Association for the Advancement of Artificial

Intelligence, 1997. AAI Technical Report SS-97-06. Available at: <http://oa.upm.es/5484/1/METHONTOLOGY_.pdf>. Access at: 20 jun. 2014.

HJORLAND, B. Fundamentals of knowledge organization. *Knowledge Organization*, v.30, n.2, p. 87-111, 2003.

GARCÍA MARCO, F. J. Ontologías y organización del conocimiento: retos y oportunidades para el profesional de la información. *El profesional de la información*, v. 16, n. 6, p. 541- 550, Nov.- Dic. 2007. Available at: <<http://www.elprofesionaldelainformacion.com/contenidos/2007/noviembre/01.pdf>>. Access at: 10 jan. 2014.

GIL, A. C. *Como elaborar projetos de pesquisa*. 4. ed. São Paulo: Atlas, 2007.

GOMES, H. E. et al. *Manual de elaboração de tesouros monolíngues*. Brasília: CNPq/PNBU, 1990.

GRUBER, T. R. Toward principles for the design of ontologies used for knowledge sharing. In: GUARINO, N.; POLI, R. (Org.). *Formal Ontology in Conceptual Analysis and Knowledge Representation*. [s.l.]: Kluwer Academic Publishers, 1993. Available at: <http://pdf.aminer.org/000/912/413/toward_principles_for_the_design_of_ontologies_used_for_knowledge.pdf>. Access at: 29 jun. 2014.

GUARINO, N. *Formal Ontology and Information Systems*. [Vilnius]: Vilnius University Faculty of Mathematics and Informatics, 1998. Available at: <<http://uosis.mif.vu.lt/~donatas/Vadovavimas/Temos/OntologiskaiTeisingasKonceptinisModeliavimas/papildoma/Guarino98-Formal%20Ontology%20and%20Information%20Systems.pdf>>. Access at: 26 jul. 2014. Amended version of a paper appeared in N. Guarino (ed.), *Formal Ontology in Information Systems*. Proceedings of FOIS'98, Trento, Italy, 6-8 June 1998. Amsterdam, IOS Press, p. 3-15.

NOY, N. F.; MCGUINNESS, D. L. *Ontology development 101: a guide to creating your first ontology*. Stanford: Stanford University, [2001?]. Available at: <<http://www.ksl.stanford.edu/people/dlm/papers/ontology101/ontology101-noy-mcguinness.html>>. Access at: 20 fev. 2014.

RAMALHO, R. A. S. *Desenvolvimento e utilização de ontologias em bibliotecas digitais: uma proposta de aplicação*. 2010. 145f. Tese (Doutorado em Ciência da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2012.

RAMALHO, R. A. S.; FUJITA, M. S. L. Aplicabilidad de ontologías en bibliotecas digitales. *Anales de Documentación*, v. 14, n. 1, 2011. Available at: <<http://revistas.um.es/analesdoc/article/view/120161>>. Access at: 12 mar. 2014.

RANGANATHAN, S. R. *As cinco leis da biblioteconomia*. Brasília: Briquet de Lemos, 2009.

SANTOS, M. T.; CORREA, R. F.; SILVEIRA, M. A. A. Estudos brasileiros sobre ontologia na Ciência da Informação. *Datagramazero*, v. 14, n. 1, 2013. Available at: <http://www.dgz.org.br/fev13/Art_05.htm>. Access at: 20 jun. 2014.

SILVEIRA, D. T.; CÓRDOVA, F. P. A pesquisa científica. In: GERHARDT, T. E.; SILVEIRA, D. T. (Orgs.). *Métodos de pesquisa*. Porto Alegre: Editora da UFRGS, 2009.

SOERGEL, D. The rise of ontologies or the reinvention of classification. *Journal of the American Society for Information Science*, New York, v. 50, n. 12, p. 1119-1120, Oct. 1999. Available at: <<http://www.dsoergel.com/cv/B70.pdf>>. Access at: 10 jan. 2014.

SVENONIUS, E. *The Intellectual Foundation of Information Organization*. Cambridge: MIT Press, 2000.

TRIVIÑOS, A. N. S. *Introdução à pesquisa em ciências sociais: a pesquisa qualitativa em educação*. São Paulo: Atlas, 1987.

USCHOLD, M.; GRUNINGER M. Ontologies: principles, methods and applications. *The Knowledge Engineering Review*, v. 11, n. 2, p. 93-136, 1996.

USCHOLD, M.; KING, M. Towards a Methodology for Building Ontologies. In: IJCAI-95: Workshop on Basic Ontological Issues in Knowledge Sharing, 1995, Montreal, Canada. *Proceedings*. 1995. Available at: <http://www1.cs.unicam.it/insegnamenti/reti_2008/Readings/Uschold95.pdf>. Access at: 20 jun. 2014.

VICKERY, B. C. Ontologies. *Journal of Information Science*, v. 23, n. 4, p. 277–286, 1997.

W3C Recomentadion. *OWL 2 Web Ontology Language Structural Specification and Functional-Style Syntax* (Second Edition). 2012. Motik, B, Patel-Schneider, P.F. and Parsia, B. eds. Available at: <<http://www.w3.org/TR/owl2-syntax/>>. Access at: 20 mar. 2015.

W3C Candidate Recomentadion. *SKOS Simple Knowledge Organization System Reference*. 2009. Available at: <<http://www.w3.org/TR/2009/CR-skos-reference-20090317/>>. Access at: Março, 2015.

ZAFALON, Z. R. Simbiose entre catalogação e recursos tecnológicos: associações entre Ciência da Informação e Linguística pelo viés dos registros bibliográficos automatizados. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 14., 2013, Florianópolis. [*Anais eletrônicos*]. 2013. Available at: <<http://enancib2013.ufsc.br/index.php/enancib2013/XIVenancib/paper/download/496/328>>. Access at: 20 maio 2014.

ZOGLAMI, K.; KERHERVÉ, B.; GERBÉ, O. Using a SKOS engine to create, share and transfer terminology data sets. In: INTERNATIONAL CONFERENCE ON SIGNAL IMAGE TECHNOLOGY & INTERNET-BASED SYSTEMS, 7th, 2011, Dijon, France. *Proceedings...* Los Alamitos, California: IEEE Computing Society, 2011. p.46-53.

ZHAO, Y; HU, C. An Ontology-based Framework for Knowledge Service in Digital Library. In: INTERNATIONAL CONFERENCE ON WIRELESS COMMUNICATIONS, NETWORKING AND MOBILE COMPUTING, 2007, Shangai. *Proceedings*. Palo Alto: IEEE, 2007. p. 5345-5348.

Academic genealogy as an approach to domain analysis



Renata Cristina Gutierrez Castanha
São Paulo State University
regutierrez@gmail.com



Maria Cláudia Cabrini Grácio
São Paulo State University
cabrini@marilia.unesp.br

Introduction

In the context of Information Science, domain is defined as a discursive or thinking community that participates in sharing of activities in which language, structure of work and cooperation patterns, forms of communication and knowledge organization, information systems and relevance criteria are reflections of studied objects belonging to that community and its role in society (HJØRLAND; ALBRECHTSEN, 1995).

Domain Analysis is a relevant approach to characterize science as it allows the identification and analysis of the conditions under which scientific knowledge is constructed and socialized; it enables observing the integration of the individual with the social context of the communities where he is inserted, and the concepts of information and knowledge have their meaning according to the shared understanding of the members of these communities (OLIVEIRA; GRÁCIO, 2013).

In this context, Hjørland (2002) suggests 11 approaches to study and acknowledge a domain, and stresses that the joint use of these approaches provides a deeper understanding of the area to be studied. Among the presented approaches, according to the author, bibliometric studies operationalize and sustain qualitative analyzes which seek to understand the socio-cognitive dimensions of science, providing consistent analysis of how knowledge is produced, organized and disseminated.

Thus, qualitative and quantitative studies of the construction of scientific intellectual spaces occupy important position in the research dimension of various fields of knowledge, making it possible to verify what is actually important or significant in a given field, so that aspects such as trends, patterns, processes, agents and their relationships can be identified and analyzed (DANUELLO, 2007).

Hjørland (2002) proposed the bibliometric approach, primarily focusing on citation studies. However, many scholars argue that, currently, the quantitative analysis of a domain stretch beyond bibliometric studies, due mainly to the emergence of new technologies for the retrieval, organization and analysis of large amounts of data (SUGIMOTO, 2014; HERRERO- SOLANA, 2001), giving rise to new research and analysis procedures, including the analysis of social networks, and more recently, in particular, the opinion of Academic Genealogy, contributing to the inclusion of cognitive aspects of information.

Analysis of the Academic Genealogy consists of studies of intellectual heritage, operationalized by interdependent relations between the students and their advisor-professors (SUGIMOTO, 2014). It consists of a specific type of social network, characterized as a tree and as scientific collaboration since their actors are the professors and students.

Considering the above, the objective of the study is, in general, to describe and understand the analysis of the Academic Genealogy approach for Domain Analysis.

More specifically, we propose to discuss the fundamental aspects of the Academic Genealogy analysis and the possible complementarity and articulation with other Domain Analysis' approaches proposed by Hjørland (2002). We also seek to present applications of the Academic Genealogy analysis for the scientific literature, as an approach to characterize a scientific domain.

This is a theoretical study that seeks to contribute to a better conceptual and methodological understanding of the Academic Genealogy's underpinnings. In addition, it is also of documentary nature as it brings examples of applications of this approach in the scientific literature, for identification and characterization of a domain.

Academic genealogy as an approach to domain analysis

The need to understand the social structures of science fostered the creation of analytical methodologies such as Social Network Analysis, which includes contributions from the sociology of science, mathematics and social history of knowledge in order to investigate the role and social relationship among the actors in a domain (BUFREM;¹⁰⁴ GABRIEL JR; SORRIBAS, 2011).

Academic Genealogy, through its tree-shape social network scheme identifies, highlights and organizes scientists and scholars, according to their mentoring relationships or supervisions, contributing to the visualization of a scientific lineage (Miyahara, 2011). It exposes a special kind of scientific collaboration in a domain, based on the contribution the advisor provides to achieve the students' research objectives, in which the student acquires new skills and tacit knowledge through this collaborative activity (BALANCIERI et al., 2004).

The ideas, contributions and points of view of a scientist can be perpetuated and continue to influence scientific thought in future generations of researchers, through the students he or she advises (ANDRAOS, 2005).

An Academic Genealogy tree can be descending or ascending. A descending tree displays, from a researcher, his/her students and the students advised by their students. On the other hand, an ascending genealogy tree, from a researcher, displays his or her supervisor and possibly the supervisor's supervisor, thus following, the opposite direction of the descending tree (MIYAHARA, 2011).

It can be said, therefore, that a descending academic genealogical study consists of a diachronic analysis of a scientific domain, and an ascending genealogical study of a synchronic analysis of a domain.

Academic genealogical analysis, a particular form of social network is based on a social paradigm, once the scientific activity presents a hierarchical social model from the relations among researchers and the way knowledge transfer and epistemic practices are consolidated through these relations. Specifically, a social hierarchical relationship of interdependence is observed between advisor and student, in which there is a particular form of knowledge dissemination and transfer, reflected in the discursive community and carried with the subjectivity existing in the area in which they are inserted.

It is understood that these mentoring practices are responsible for the tacit and explicit transfer of knowledge during the interdependent interactions between advisor and student (GIRVES; WEMMERUS, 1988).

In this sense, Academic Genealogy is characterized as a fundamental approach to study the knowledge transfer in a domain, providing a means to measure and analyze these interdependent interactions in the academic sphere. Although the life of the scholar is finite, his/her scientific contributions are impacted by successive generations of mentees (SUGIMOTO, 2014).¹⁰⁵

Thus, Academic Genealogy trees and forests, through mentoring relationships, contribute to the visualization and characterization of groups and active scientific currents in a domain.

In this context, it is considered that genealogical studies, associated with historical and epistemological studies proposed by Hjørland (2002), allow a broad, objective and contextual view of a scientific field.

It is also noted that Academic Genealogy studies are shown significant for assessing the impact of academic mentoring in the scientific development of a community and in identifying the main actors (ROSSI, MENA-CHALCO, 2014).

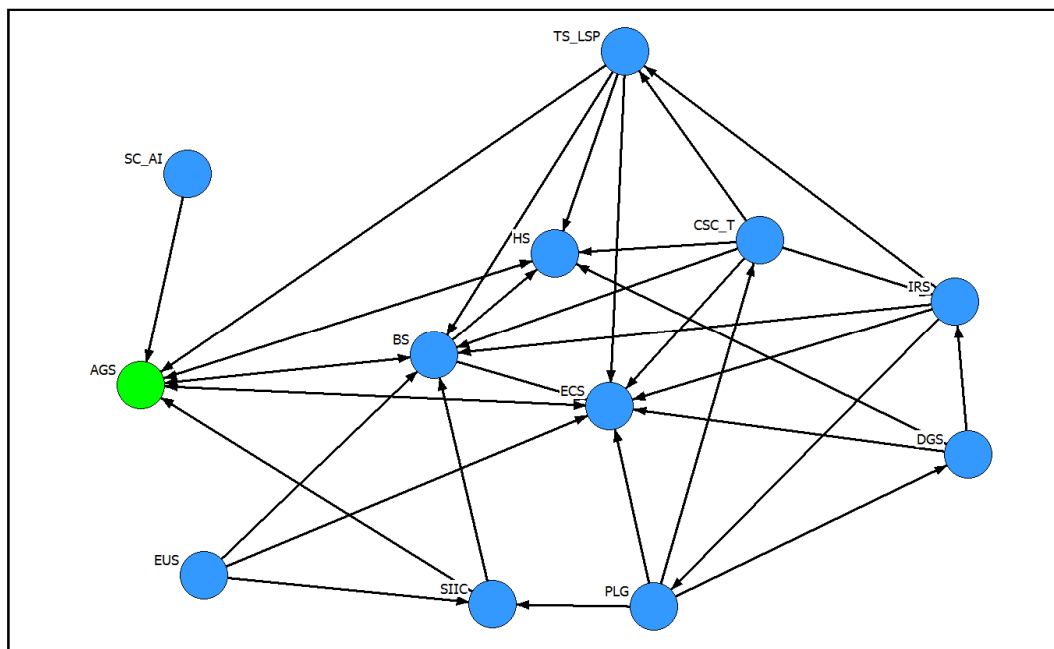
In this context, these studies, articulated to bibliometric studies proposed by Hjørland (2002), provide an important contribution to Domain Analysis by allowing to know and to characterize trends, patterns, processes, dominant thoughts, important actors and their relationships.

In order to introduce Academic Genealogy studies as an approach to Domain Analysis, we elaborated a diagram (Figure 1) of the 11 approaches proposed by Hjørland (2002), plus Academic Genealogy studies, with the associations proposed by the author, represented by vectors, plus the associations considered to genealogical studies.

Firstly, it is observed that the historical and epistemological and critical approaches, although essential to most other approaches, they do not require, according to Hjørland (2002), complementary approaches. The author considers that the historical approach is a substantial method to understand a domain, to offer a deep and cohesive perspective, and the epistemological and the critical approaches provide knowledge on the theoretical foundations and the evolution of the area being studied.

In addition, Hjørland (2002) does not associate the approach on scientific cognition, specialized knowledge or artificial intelligence to any other type of proposed approach.

Figure 1 - Diagram associating 11 approaches to Domain Analysis by Hjørland (2002), plus Academic Genealogy Studies



Caption: AGS = Academic Genealogy Studies; PLG = Producing literature guides or subject gateways; CSC_T = Constructing special classifications and thesauri; IRS = Indexing and retrieving specialties; EUS = Empirical user studies; BS = Bibliometrical studies; HS = Historical studies; DGS = Document and genre studies; ECS = Epistemological and critical studies; TS_LSP = Terminological studies, language for special purpose, database semantics and discourse studies; SIIC = Structures and institutions in scientific communication; SC_AI = Scientific cognition, expert knowledge and artificial intelligence (AI).

Source: by the authors

Regarding the Academic Genealogy studies approach, it is considered that its analysis can be enhanced when combined with historical, bibliometric and epistemological and critical studies, contributing to the consolidation of meaningful understanding of the studied domain.

On the other hand, it is also considered that bibliometric studies, historical studies, epistemological and critical studies, terminology studies, the studies of structures and institutions and also the studies of cognition and specialized knowledge can benefit when combined with Academic Genealogy studies, enriching the understanding of the origins,

evolution and dissemination of knowledge in the domain and contributing to the understanding of currents of thought and trends.

Several studies analyzing different scientific domains have used the Academic Genealogy analysis approach, namely: Mathematics (CHANG, 2011; MALMGREN et al, 2010; ROSSI, MENA-CHALCO, 2014.), Information Science (MARCHIONINI et al. 2007;¹⁰⁷ RUSSELL; SUGIMOTO, 2009; SUGIMOTO et al, 2011), Neuroscience (DAVID; HAYDEN, 2012), among others. It is observed that they are very recent studies and, mostly carried out in order to analyze a domain, identifying how the discursive community recognizes its theoretical reference. We highlight, however, that in these studies, the theoretical bases of domain analysis were not explained, i.e., these researchers, even performing domain analysis, do not explicit or are aware that they are performing domain analysis.

It should be noted, though, that much of what have been published in the scientific literature on Academic Genealogy, has shown that its features are based on graphs and social networks, since Academic Genealogy is a particular type of network, hierarchical and defined as a tree (family).

Conclusions

This study aimed to contribute to a better conceptual and methodological understanding of Academic Genealogy Studies involving the analysis of interdependent hierarchical relationship between advisor and student, and highlights the contribution of these studies as an approach to Domain Analysis, which, in cooperation with other approaches proposed by Hjørland (2002), make it possible to know and characterize a scientific domain.

It is considered that, by showing the origins, evolution and dissemination of knowledge in a scientific domain, through the supervision and mentoring relations, we contribute to the understanding of group formation, schools of thought and trends in a domain, Academic Genealogy Studies are relevant Domain Analysis approach.

In this sense, it considers that Academic Genealogy Studies can also be of important contribution to the study of the sociology of science, by allowing a synchronic

and diachronic visualization of the relations among researchers and scholars from a scientific domain.

Finally, it is considered that analysis of Academic Genealogy can consist of a significant analytical approach to the study of scientific domains, and combined with quantitative analysis, such as bibliometric studies and qualitative analysis, such as¹⁰⁸ historical, epistemological and critical studies, for better characterization and understanding of a particular domain.

References

The references was made following the ABNT rules.

ANDRAOS, J. Scientific genealogies of physical and mechanistic organic chemists. *Canadian Journal of Chemistry*, v.83, n.9, p.1400–1414, 2005.

BALANCIERI, R.; BOVO, A.; KERN, V.; PACHECO, R. C.; BARCIA, R. M. A análise de redes de colaboração científica sob as novas tecnologias de informação e comunicação: um estudo na Plataforma Lattes. *Ciência da Informação*, v.34, n.1, 2005.

BUFREM, L., GABRIEL JÚNIOR, R., SORRIBAS, T. Redes sociais na pesquisa científica da área de ciência da informação. *DataGramaZero*, v.12, n.3, 2011.

CHANG, S. *Academic genealogy of mathematics*. Singapore: World Scientific Publishing, 2011.

DANUELLO, J. *Produção científica docente em tratamento temático da informação no Brasil: uma abordagem métrica como subsídio para a análise do domínio*. 122 p. Dissertação (mestrado) – Universidade Estadual Paulista, Faculdade de Filosofia e Ciências, 2007.

DAVID, S.; HAYDEN, B. Neurotree: A collaborative, graphical database of the academic genealogy of neuroscience. *PLoS ONE*, v.7, n.10, 2012.

GIRVES, J.; WEMMERUS, V. Developing models of graduate student degree progress. *Journal of Higher Education*, v.59, n.2, 163–189, 1988.

HERRERO-SOLANA, V. Producción científica de la Universidad Nacional de Mar del Plata: análisis de dominio. *Nexos*, v.8, n.14, p. 4–10, 2001.

HJØRLAND, B. Domain analysis in information science: Eleven approaches – traditional as well as innovative. *Journal of Documentation*, v.58, n.4, p.422–462, 2002.

HJØRLAND, B.; ALBRECHTSEN, H. Toward a new horizon in information science: Domain-analysis. *Journal of American Society Information Science*, v.46, p.400–425, 1995.

MALMGREN, R.; OTTINO, J.; AMARAL, L. A. The role of mentorship in protégé performance. *Nature*, v.465, n.3, p.622–627, 2010.

MARCHIONINI, G.; SOLOMON, P.; DAVIS, C.; RUSSELL, T. Information and library Science MPACT: a preliminary analysis. *Library & Information Science Research*, v.28, n.4, p.480–500, 2006.

MIYAHARA, E.K. *Genealogia Acadêmica Lattes*. Monografia (Bacharelado em Ciência da Computação) – Universidade de São Paulo, Instituto de Matemática e Estatística, 2011.

| 109

OLIVEIRA, E. T; GRACIO, M. C. C. Studies of Author Cocitation Analysis: a bibliometric approach for domain analysis. *Iris*, v.2, n.1, p.12-23, 2013.

ROSSI, L.; MENA-CHALCO, J. Caracterização de árvores de genealogia acadêmica por meio de métricas em grafos. In: *Brazilian Workshop on Social Network Analysis and Mining*, Brasília-Brazil, p.1–12, 2014.

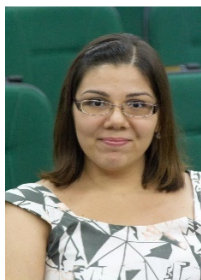
RUSSELL , T.; SUGIMOTO, C. R. MPACT family trees: Quantifying academic genealogy in library and information science. *Journal of Education for Library and Information Science*, v.50, n.4, p.248–262, 2009.

SMIRAGLIA, R. ISKO 11's Diverse Bookshelf: an editorial. *Knowledge Organization*, v. 38, n.3, p.179–186, 2011.

SUGIMOTO, C.R. Academic Genealogy. In: CRONIN, B.; SUGIMOTO, C.R.(Eds.). *Beyond bibliometrics: harnessing multidimensional indicators at scholarly impact*. Cambridge: MIT Press, 2014.

SUGIMOTO, C. R.; NI , C.; RUSSELL , T. G.; BYCHOWSKI, B. Academic genealogy as an indicator of interdisciplinarity: An examination of dissertation networks in library and information science. *Journal of the American Society for Information Science and Technology*, v.62, n.9, p.1808–1828, 2011.

The epistemological dimension of Document Analysis of fictional content in Knowledge Organization



Roberta Caroline Vesú Alves
São Paulo State University
robertavesu@gmail.com



João Batista Ernesto de Moraes
São Paulo State University
jota@marilia.unesp.br

L10

Introduction

The interdisciplinary nature of Document Analysis in narrative fictional texts includes some theoretical aspects of Discourse Analysis and Greimas' Semiotics, which underpin their analysis procedures, synthesis and representation (FIORIN, 2008; GREIMAS 1971; GUIMARÃES; MORAES; GUARIDO , 2007).

These epistemological aspects incorporated into Document Analysis collaborate to its theoretical and methodological basis and develops areas maintaining theoretical dialogue, such as Thematic Treatment of Information, Documentary Linguistics, Documentary Text Linguistics and Knowledge Organization.

Thus, the objective of this study was to analyze interdisciplinary in Document Analysis in fictional narrative texts, studied by researchers at Universidade Estadual Paulista (UNESP), Marília-SP-Brazil, verifying their epistemological aspects, also related to Knowledge Organization. Hence, we used exploratory interdisciplinary methodology in order to obtain qualitative analysis of information.

It was found that Document Analysis for these works presents interdisciplinary aspects that support its procedures, characterized primarily as Greimas' Semiotics and Linguistics within the text and speech, which also epistemologically enrich other areas, as follows.

Document Analysis and Knowledge Organization

According to Gardin et al (1981), Document Analysis is concerned with the representation procedures of documents in a distinct way from the original ones, through classification, translation, abstracting and indexing, with the objective of facilitating retrieval.

| 111

For Guimarães, Moraes, Guarido (2007), Document Analysis develops procedures of analysis, synthesis and representation of documents for retrieval purposes, through the analytical steps (procedures of reading the important parts of the document, with metacognitive strategies for identification of concepts) and synthetic steps (selection of concepts, condensation in abstracts and representation in indexes).

Moreover, according to Guimarães (2008, 2009), it consists of theoretical current of Thematic Treatment of Information, of French origin, which began in the late 1960s with the studies of Gardin and Coyaud, also called Document Analysis (*analyse documentaire*). According to Guimarães et al (2012), Document Analysis in France and in Brazil designates the Thematic Treatment of Information, while in Spain is included in the Thematic Treatment (description of form and content).

Guimarães et al (2012, p.197) claims that Document Analysis

[...] has found fertile ground for development, particularly in Brazil and Spain, from the 1980s: in Brazil, through theoretical studies by Group Temma, with strong influence from Gardin, led by Johanna Smit, and in Spain from a strong connection of authors such as López Yepes and Sagredo Fernández in Madrid, with Otlet's tradition or still with Nuria Amat in Barcelona, with Chaumier's ideas.

Thematic Treatment of Information, according to Guimarães (2008), contains two more theoretical currents in its epistemological basis, subject cataloguing (north American) and indexing of English origin, contributing to its theoretical and methodological consolidation.

For example, in the United States and England, the studies were focused on information products and tools, and in France there was the "[...] concern with the development of theoretical and methodological framework for the Thematic Treatment of Information itself", as, until then the used procedures were guided only by common sense (GUIMARÃES, 2009, p.109).

In this context, Thematic Treatment of Information studies the description and representation of form and content of documents for retrieval, is inserted in Treatment of information, a subarea of Library and Documentation, which maintains theoretical dialogue with Knowledge Organization (BARITÉ, 2001).

According to Barité (2001), Knowledge Organization is concerned with the development of techniques of construction, management, use and evaluation for documentary languages through assumptions that determine the knowledge socially established from information and transformed into information when socialized, requiring organization for different uses, among other aspects (BARITÉ, 2001).¹¹²

In this sense, Esteban Navarro and García Marco (1995, p.149, our translation), state that the field of Knowledge Organization consists of,

[...] a discipline devoted to the study and development of fundamentals and techniques of planning, construction, management, use and evaluation of systems of description, cataloging, ordering, classification, storage, communication and retrieval of documents created by man to testify, preserve and transmit his knowledge and his actions, from their content, in order to ensure its conversion into information able to generate new knowledge. Therefore it is a three-dimensional science, which concerns principles, methods and tools put into action for the management of human knowledge from a triple perspective: its representation, its organization and its documentary communication.

Thus, it is verified that Knowledge Organization is concerned with organizational and representation aspects needed for knowledge socialization also materialized in documents, among other aspects. However, underlying representation, are the analysis and synthesis procedures of document content, which involves studies of all steps of Document Analysis in Knowledge Organization.

Interdisciplinary epistemological aspects of Document Analysis of content for fictional narrative texts

The theoretical and methodological interdisciplinary aspects of Document Analysis of Content procedures, to Guimarães (2009), include areas such as Linguistics, Terminology, Logics, Cognitive Psychology, among others. In this context, the interdisciplinary theoretical and methodological aspects that epistemologically support

Document Analysis of narrative fictional texts in studies by researchers from UNESP Marilia-SP-Brazil were verified.

These studies use the application of Generative Trajectory of Meaning for identification of aboutness of these texts, as Moraes (2011, 2012a, 2012b), Moraes and Guimarães (2006), Guimarães, Moraes and Guarido (2007), among others. Thus, they use the Generative Trajectory of Meaning as a metacognitive strategy of identifying important parts of the text for thematic content analysis. ¹¹³

Generative Trajectory of Meaning consists of a theoretical model for understanding different internal structural levels of verbal and non-verbal texts and their layers of meaning (syntax and semantic aspects), such as narrative and scientific due to the inherent narrative. In addition, it belongs to the area of Discourse Analysis and Greimas' Semiotics, it is based on Greimas's postulates and Propp's, Hjelmslev's studies, among others, therefore, with structuralist characteristics, due to the interrelations between the abstract structures of the text and their meanings (GREIMAS, 1971; FIORIN, 2008).

The Generative Course presents the deep and fundamental structures, narrative (manipulation, competence, performance and penalty) and discursive (themes and figures), and the semantic aspects understood by isotopy (FIORIN, 2008). Therefore it allows identification of aboutness or main theme, which is presented in deep structure (MORAES, 2011, 2012a, 2012b).

We found, therefore, that this theoretical model allowed the identification of aboutness from the text rules in Document Analysis (HUTCHINS, 1977; BEGHTOL, 1986; MORAES, 2011, 2012a, 2012b). Also, it provided the classification of themes of fictional narrative text for construction purposes of indexes and index concepts (GUIMARÃES; MORAES; GUARIDO, 2007).

Moreover, the Generative Trajectory of Meaning operates in Document Analysis, as well as Van Dijk's Text Linguistics (2000) sets out the structure parameters and text semantics, through microstructures, macrostructures, textual superstructures and macro-rules, thus, demonstrating important textual aspects for analysis, synthesis and representation.

The interdisciplinary of Document Analysis for fictional narrative texts is based on the theoretical model of Generative Trajectory of Meaning, which theoretically and

methodologically characterizes with Greimas' Semiotics and linguistic aspects within the text and discourse. Therefore, it also extends the theoretical and epistemological perspectives of Thematic Treatment of Information and Knowledge Organization through interdisciplinary with these areas.

It is also worth noting that studies in Document Analysis focused on the text also collaborate with the development of other subfields of Knowledge Organization, such as Documentary Linguistics and Documentary Text, for they are disciplines or areas concerned with the study of textual structures and their meanings, applied for information organization and representation.

Documentary Linguistics, originally proposed by García Gutiérrez, according to Lara and Tálamo (2007), joints the studies on language issues related to documentary languages in Brazil. Also, it contains the branch of documentary semiotics, Peirce's studies applied in the communication of documentary languages. But, according to García Gutiérrez (1998, our translation)

[...] we must emphasize the importance of consolidation and expansion of Documentary Linguistics as a discipline that deepens its foundations in the postulates of the sciences of language, semantics and grammar, mainly before intersections occurred with related fields, such as Discourse Analysis, Content Analysis and, in general, with the so-called, and in construction, Cognitive Sciences.

Thus, we verify that the studies within the text focused on analysis, synthesis and representation also constitute this area, which allows to include studies on Generative Trajectory of Meaning in Documentary Linguistics.

The Documentary Text, in turn, studies the discursive and textual aspects of organization and representation of content, related to the branch of documentary morphology content and the documentary form of content (IZQUIERDO ALONSO, 2004).

The studies with Hjelmslev's focus explore the general principles of internal structures of messages, the form and content of the texts, as well as the relations and function of their structural elements, among other aspects (IZQUIERDO ALONSO, 2004). However, they are also studies on the structure and meaning of texts, therefore, they are parallel and complementary to other linguistic theories and semiotics alike, which allow structural layers and text sense.

Given the above, it was found that the Generative Trajectory of Meaning has Greimas' linguistic and semiotic character in the text and discourse and also contributes to the epistemological development of Document Analysis and other related fields.

Conclusion

The interdisciplinary used by Document Analysis of fictional narrative texts, studied by researchers from UNESP Marilia-SP-Brazil, is focused on the theoretical approach of Generative Trajectory of Meaning, from Discourse Analysis and Greimas' Semiotics. Also, it is constituted with theoretical, methodological and epistemological aspects that support Document Analysis in analysis, synthesis and representation procedures of content of these texts.

Generative Trajectory of Meaning allows the identification of aboutness from significance rules and text structure to construction indexes and indexing concepts, conceptually and methodologically enriching Document Analysis and areas maintaining theoretical dialogue.

In this sense, it was found that these interdisciplinary theoretical aspects have Greimas' linguistic and semiotic character in text and discourse, and collaborate to epistemologically support areas developed with studies of Document Analysis such as the Thematic Treatment of Information, Documentary Linguistics, Documentary Text and the field of Knowledge Organization.

Indexing languages and domain analysis



Vera Lucia Ribeiro Guim
São Paulo State University
verinha.guim@gmail.com



Mariângela Spotti Lopes Fujita
São Paulo State University
fujita@marilia.unesp.br

16

Introduction

One of the treatments information science addresses is information retrieval and representation. Thus, using clear and objective language designated by a professional indexer is necessary. These languages are named indexing languages.

Indexing languages are artificially constructed languages that aim to succinctly represent the content of documents. The language as a working object brings up the need to study sciences that also seek in language their objects of study, such as linguistics, among others.

In order to contextualize this issue, we focus on domain analysis, analyze which comprehends historical, cultural and linguistic issues of knowledge or user community, about the representation of information on systems and information environments.

Approached in many studies, domain analysis has been the subject of several studies in order to further support its proposed themes. Given these considerations, in this study, we study carry out a literature review and search, based on theories of domain analysis, the theoretical underpinnings that contribute and meet indexing languages, their dynamism and representation, within information science.

Indexing Languages

Embedded in Information Science, understanding indexing language emerges as an essential tool for information organization and production. Thus language is contextualized within documentary treatment as an activity of information units whose intension is to meet the goals and objectives of the organization.

NP 4036 norm defines indexing language as the controlled set of terms chosen in a natural language and used to represent, in a coordinated manner, the content of the documents.

Indexing languages are auxiliary tools in activities of information representation and retrieval. They act on information systems guiding the indexer about the best terms to represent the subjects of a document and guiding researchers in developing information search strategies in the system. (MOURA, 2002, p.4)

Libraries or information units make use of document or indexing languages to describe the contents of document collections in order to store them and retrieve the information in the documents. Therefore, documentary languages are divided into two classes, hierarchical documentary languages and alphabetic documentary languages.

Hierarchical documentary languages or pre-coordinated are notational, that is, the terms coordinate before their determination, examples of hierarchical documentary languages are classificatory systems such as UDC (Universal Decimal Classification) and CDD (Dewey Decimal Classification).

In alphabetic documentary languages or post-coordinated, the terms coordinate themselves after their determination. And they are formed by controlled vocabularies such as thesauri or subject headings.

For an optimal use of documentary languages, it becomes necessary to connect natural language used by the user with the language designated by the retrieval system, once in certain regions of the country one searched word may have different meanings. Thus, in order to facilitate the communication process between the user and the system, control instruments are structured for the terms to be used, therefore, a full mastery in the language used is necessary so that the terms can be known and easily used by the indexer of the system and consequently by the user.

According to Vale (1987), the choice for an indexing language is an influencing factor for the effectiveness of an information retrieval system, thus such a choice needs

to consider the system's objectives, the user's profile and the scope or specificity of the subject being treated.

Pre-coordinated *and* *Post-coordinated* *languages* _____
| 118

By organizing the information following a hierarchical and systematic structure, classifications are considered pre-coordinated languages, as the concepts represented by the indexer at the moment of the representation, "are instruments par excellence used in knowledge organization, first in physical arrangement, then, and in parallel, logic and systematic organization of knowledge in catalogs "(SIMÕES, 2008, p.36, our translation).

The structure offered by classifications allows information to be grouped coherently, contextualized and logically.

The post-coordinate languages, according to Lancaster (2002, p. 22) seek to ,

1. Facilitate the consistent representation of subjects by the indexers and users that retrieve, avoiding the dispersion of the related elements.
2. Facilitate conducting extensive research on a subject linking terms with paradigmatic or syntagmatic relations.

The subject heading lists and thesauri continue to be the most used post-coordinated controlled languages in information representation and retrieval.

DDC and UDC

The most popularized classification systems date from the nineteenth century: Dewey Decimal Classification (DDC) and Universal Decimal Classification (UDC). These bibliographic classification schemes seek to map knowledge in different areas featuring subjects based on hierarchical divisions from the general to the most specific.

For Cintra (2002, p.44, our translation)

In the bibliographic classification systems, the hierarchical structure is given by notation (decimal, in the case of DDC and UDC). The vertex of the hierarchical chains is constituted by conventional disciplines that are subdivided successively. The subjects are indicated by numerical or alphanumeric notation, depending on the type of system.

Currently, in its twenty-second edition, DDC is considered the most widely used classification system in the world, adopted in more than 135 countries and translated into more than thirty languages.

The Universal Decimal Classification (UDC) was created in 1905, built by the Belgians Paul Otlet and Henri La Fontaine. The authors adapted UDC system crediting a more enumerative and linear character to it, which led to an improvement in the representation of subjects, whether simpler or more complex, through mechanisms of combination.¹¹⁹

Thesaurus

For UNESCO (1973, p.6) the thesaurus,

Is a controlled and dynamic vocabulary of semantically and generically related terms which comprehensively covers a specific domain of knowledge [...] it is a terminological control device used to translate from the natural language of documents, indexers, or users into a more constrained "system language" (documentation language, information language), as well as to translate the system language back into natural language.

According to their main users, thesauri are suited to specialized areas, in general they are easier to use, require less time from the indexer and were faster for information retrieval. Therefore, they follow the post-coordination principles.

According to IBICT (1984, p. 1-2, our translation), the main purposes of a thesaurus are:

- a) to control the terms used for indexing through an instrument that translates the natural language of the authors, users and indexers into a more controlled language;
- b) to uniform, before this documentary language, the professional indexing procedures into an institution or into a cooperative network;
- c) to limit the number of terms required to explicit the concepts exposed by the authors of an area;
- d) to assist the information retrieval task by providing appropriate terms for the search strategy.

The thesauri, examples of controlled vocabulary, carry, as its main function, the terminological vocabulary control used in a specific area of knowledge.

Subject Headings

Subject headings seek to represent the subjects in structured headings. However, they are considered one of the most costly languages, as they require the indexer a long to perform, however, the specificity and accuracy of information are more assured, allowing, in addition, a more precise description of complex subjects, and allow a more immediate perception by the user.

Subject headings represent a very structured and pre-coordinated language, giving the researcher limited possibilities to modulate his research. However, today there are vocabularies called subject heading lists and are used in post-coordinate systems, such as MESH (Medical Subject Headings). (CESARINO; PINTO, 1978, our translation)

Thus, such as every language, subject heading lists also require updates since the greater the documentation volume the higher is the degree of specificity and this requires an adequate control and quality seeking to benefit the indexer, the institution and the user of this type of language.

Documentary languages are necessary, since they seek to reconcile the vocabulary to be used by the system, by the indexer and by the users so that different terms are used for identical subjects thus avoiding similar documents end up separated.

Therefore, documentary languages end up contributing to a delimitation of the domain of fields of study. Cintra et.al. (2002) state that "this is possible as for each preferred unit integrated in a documentary language, it is necessary that a concept or notion is matched [...] since only the notional organization of an area allows the use of effective tools for treatment and information retrieval."

In order to assess the language used by retrieval systems, we seek to analyze which domain the language belongs to, in order to verify whether the language is pre- or post-coordinated in addition to its representation and preparation.

Domain Analysis

Domain analysis seeks to identify and analyze specific areas in order to build an information system, also seeking to build information representation models according to the specific fields of knowledge and their features.

To Danuello (2007), domain analysis "is a process by which one can verify what is important or meaningful about any field of human endeavor, identifying elements that allow analyzing a scientific context." Lorenzon (2011) also states that it "is based on historical, cultural and linguistic conditions of discursive communities (users) that direct the construction of thesauri and classification systems for information representation in their information environments." ¹²¹

Because it is not possible to predict the information needs of users, and to be aware that the indexer can come across various fields of knowledge, according to Hjørland (2002, p.422) domain analysis turns to the study and structuring of subject areas, likewise, to classification systems, however, with the social contextualization of concepts, reflecting the criteria that a document should never be considered in isolation.

Therefore, in order to analyze a domain, we seek to identify, explain and predict all the facts that may occur in a given set of terms within that domain. This happens so that they may serve as a source of reference when ambiguities arise, as a common knowledge repository, directly helping in communication, in learning and in reusing these terms and concepts at any level of abstraction.

Hjørland (2002) determined that for knowledge organization, it is necessary that the representation be made considering social, cultural, linguistic and cognitive knowledge of certain specific areas of knowledge.

Authors such as Hjørland and Tennis, among others, discuss domain analysis and their theories have greatly contributed to the consolidation of this analysis model in Information Science.

Thus, we believe that the methodological theoretical contribution of this approach makes it possible to contextualize the search within specific fields of indexing languages.

Indexing Languages and Domain Analysis

Thus, Hjørland (2002, p.430) concludes that indexing and information retrieval are always specific, and recommends that researchers and professionals in the area of indexing and retrieval meet the demands of different domains in organization and retrieval systems highlighting that a stronger focus on different fields can make our area more realistic and relevant in different environments or contexts. ¹²²

Regardless of the language chosen by the institution at the time the indexer decides which language to use, there is a starting point for its creation and basic research and for that, it is necessary to turn the look into the conception of domain analysis that seeks to follow the methods of social sciences, according to their actors and external collective to the subjective attitudes.

A different form of domain analysis arises from the need to understand a domain to construct or revise an information system. For example, the editors of the Dewey Decimal Classification revise it in order to address changes in the domain represented by this universal classification. (TENNIS, 2012)

Such as for research, the construction of an information system is based on domain analysis, as this,

[...] is a methodology for constructing models of information representation based on the investigation of certain characteristics of specific knowledge domains: the identification of cultural, historical and linguistic conditions that impose particular requirements for the construction of domain models such as classification systems or thesauri. It also includes an epistemological point of view to identify scientific and technical paradigms, research approaches and knowledge interests in the domains covered. (NOVELLINO, 1996)

Thus it is clear that different languages and methodologies and methods translate themselves while domains once they allow the construction of systems and research in view of their objectives and fundamentals.

Considerations

Before various applications and adaptations, studies on domain analysis are

increasingly interrelated and interconnected with the most different areas of knowledge. Their relevance in the face of languages is settled and inferred in their studies.

The functions and actions of domain analysis for indexing is of clear importance and conditionality. Each application points its foundation and approach as well as its length and exclusion in order to contribute to the deepening of representation and translation issues and value of domain analysis, considering its reach and scope in relation to indexing languages.¹²³

Considered fundamental for a proper description and indexing to systems and informational means, it is assumed that the experiences that indexing languages are affected may suggest the culture, the environment, education and even emotional that domain analysis uses for research purposes.

We also infer that structured languages, their translations and representations in information units directly influence the organization's purposes.

Considering this, the authors cited in this study reinforce the need for a structured indexing language from the theoretical depth provided by domain analysis and which seeks to be understood as an indispensable tool to promote and add value, effectiveness and consistency to information.

References

The references was made following the ABNT rules.

CESARINO, M.A. N; PINTO, M.C.M.F. Cabeçalho de assunto como linguagem de indexação. *R. Esc. Bibliotecon.* Belo Horizonte, v.7, n. 2, p.268-88, Set. 1978.

CINTRA, A.M. M et.al. *Para entender as linguagens documentárias*. 2. ed. rev.e ampl. - São Paulo: Polis, 2002

DANUELLO, J. C. *Produção científica docente em tratamento temático da informação no Brasil: uma abordagem métrica como subsídio para a análise de domínio*. Marília: UNESP, 2007. Dissertação (Mestrado) – Programa de Pós-Graduação em Ciência da Informação - Universidade Estadual Paulista (UNESP).

HJØRLAND, B. Domain analysis in information science: Eleven approaches – traditional as well as innovative. *Journal of Documentation*, London, v.58, p.422-462, 2002. Available at: www.db.dk/bn/publikationer/Filer/JDOC_2002_Eleven_approaches.pdf

IBICT. *Diretrizes para elaboração de tesouros monolíngues*. Brasília, 1984.

LANCASTER, F.W. *El Control del Vocabulario en la Recuperación de Información*. 2. ed. Valencia, 2002.

LORENZON, E.J. *Análise de Domínio para Avaliação de Tesouros: Uma experiência com a cadeia produtiva do calçado no Brasil*. Marília: UNESP, 2011. Tese (Doutorado) – Programa de Pós-Graduação em Ciência da Informação - Universidade Estadual Paulista (UNESP).

| 124

MOURA, M.A.; SILVA, A.P; AMORIM, V.R. A concepção e o uso das linguagens de indexação face às contribuições da semiótica e da semiologia. *Informação e Sociedade: Estudos*, v.12, n.1, 2002

NOVELLINO, M.S.F. Instrumentos e metodologias de representação da informação. *Informação e Informação*, Londrina, v.1, n. 2, p. 37-45, jul./dez. 1996.

SIMÕES, Maria da Graça. *Da abstração à complexidade formal: relações conceptuais num tesouro*. Coimbra: Almedina, 2008

TENNIS, J.T. Two axes of domains for domain analysis. *Knowledge Organization*, v. 30, n. 3/4, 2003, p.191

TENNIS, J.T. Com o que uma análise de domínio se parece no tocante a sua forma, função e gênero? *BRAJIS*, Marília, v.6, n.1, p.3-15, jan./jun. 2012.

UNESCO. *Guidelines for the establishment and development of monolingual thesauri*. Paris, 1973.

VALE, Eurides A. do. Linguagens de indexação. In: SMIT, Johanna W. (Coord.). *Análise documentária: a análise da síntese*. Brasília: NCT – CNPQ, 1987. Cap. 1, p.12 – 26.

Reader-Interest classification as a method of stock control: The McClellan Legacy



Daniel Martínez-Ávila
São Paulo State University
dmartinezavila@marilia.unesp.br



M P Satija
Guru Nanak Dev University
satija_mp@yahoo.com

L25

Introduction

A "reader-interest classification" is "a classification designed to serve the immediate needs of the targeted users. Such systems may violate the filiation sequence to bring together disparaged subjects needed by a user group. Indeed these are useful in mission oriented or multidisciplinary subjects. In a commerce college, e.g., it may be more pragmatic to place commercial law with commerce at 380. It is true to say that reader's interest classification adopted so far are not satisfactory in the long run and sometimes correspond to ephemeral vogues. It reflects a middle level of ambition in knowledge organisation. It is a compromise between ad hoc classification and rigorously scientific classification" (Satija 2004, p.182). Although not exempt from controversy, the study of reader-interest classifications has been recently reviewed (e.g., Satija 2009) and revived in the light of the recent cases of adoption of the book industry classification BISAC (e.g., Martínez-Ávila and San Segundo 2013; Martínez-Ávila et al. 2014). In this paper, we review the classic literature on reader-interest classifications to reveal one of the key and most remarkable claims behind these cases: the use of reader-interest classification as a method of stock control.

Reader-interest classifications and stock control

When dealing with physical collections (as did all cases which this work has examined in the literature) one of the main advantages/characteristics of reader-interest classifications was their relationship with stock control, stock management, and arrangement. Stock control, generally speaking, is the process of maintaining an appropriate level of stock to meet the demand while keeping the costs to a minimum.¹²⁶ Stock control was claimed to be one the main goals of many reader-interest classification projects such as in libraries of Cambridgeshire (Chandler 1982), Cheshire (Astin 1982), East Sussex (Donbroski 1982) and Surrey (Betts 1982). All these libraries claimed to implement the reader-interest classifications for a variety of purposes indirectly related to stock control. In Cambridgeshire the reader-interest classification was claimed to be adopted to improve self service, in East Sussex for the development of the collection, and in Surrey as a new policy of provision of stock to identify the demand. In all these cases stock control was considered a vital part of the projects and therefore linked to a variety of other aspects and advantages. According to Ainley and Totterdell, "the relationship of stock control to categorisation is vital - in fact the two are really inseparable" (Ainley & Totterdell 1982, p.126). It might be added that they are inseparable because they are mutually interdependent: reader classifications affected stock control, and stock control was usually one of the main advantages/desired goals of reader-interest classifications.

In any case, the relationship between these two elements seems based on the assumption that reader-interest classifications not only meet user's needs, but enable the most relevant areas of the collection to be identified. Once these areas of stock are physically isolated, it is then possible to identify reader's interests according to the demand/supply relation. This use of reader-interest classification allows better policies of book selection, arrangement, and weeding, while keeping the most adequate levels of stock for the user. In addition, the relationship of demand/supply and the readers' interest was also found to be linked to the concept of stock management.

Stock management, as defined by Astin, is "the committed operation of a system of procedures for the informed allocation and redistribution of stock controlled by data as to potential and actual use against predetermined targets" (Astin 1982, p.15). Even when not serving a primary purpose, stock management was claimed to be an advantage of reader-interest classifications -as was discovered with the first reader-interest

classification put into practice at Detroit Public Library. Observing the effects of this new system upon the collection, Rutzen declared that "[the reader-interest classification] affects book selection. We are beginning to be aware that some books generally purchased by branches operating under the old system may not be included by those using this classification, and vice versa" (Rutzen 1952, p.482). | 127

Other libraries adopting a reader-interest classification also noticed the practical benefits of physical display or visualization. Andrew Miller, at Glasgow City Libraries, claimed that "arranging the stock by categories does increase demand and highlights deficiencies in stock holdings" (Miller 1992, p.132). Sapiie pointed out that: "it was found that the reader-interest classification was also useful for the meaningful collection of data and assessing and editing stock because it provides a visual means to determine the size of various collections, to identify low use or surplus stock and areas for future stock provision. The reorganization process provides an excellent opportunity for stock weeding. Since the stock is physically divided into manageable groups, a reader-interest classification provides an easy way to determine how many books of a particular genre a library owns. Future stock selection is facilitated because reader-interest classification gives a clear indication of reading preferences" (Sapiie 1995, p.151).

The concept of stock management was closely related to the control of stock levels, shelf stock and the proper number of books needed for reader-interest classifications. For Douglas Betts, librarian at Surrey, reader-interest classifications were linked to the adequate control of stock levels: "some form of categorisation is an important pre-requisite for constructive thinking about stock levels" (Betts 1982, p.70). Yet at the same time it was again difficult to link stock control solely to the classification system; at Surrey, stock reduction was a basic requirement for many measures related to guiding and space, such as "face-on" and integration of all sizes books. Importantly, stock reduction was also related to the need to maximize resources and loans. As pointed out by Alex Wilson (and later by Ainley and Totterdell) a shelf arrangement based on reader interest categories should be "a first call on limited resources" (Wilson 1978). The point was to maintain stock in certain areas at the correct levels -arguably at a minimum- in order to serve a significant number of users and to thus to save costs and to avoid distraction to the reader. After determining which areas of interest (or categories in the reader-interest classification

schemes) these should be, the collection was to then focus on developing these areas in order to improve the indicators of library performance and to save readers from information overload.

The McClellan formula and reader-interest classifications

The principles of these methods of stock control, in relation to reader-interest classifications, were first developed by Archibald William McClellan, Chief Librarian of Tottenham, who devised a system to determine some kind of proportional representation of stock based on a regular (3-4 times a year) count of books on shelves and on loan. McClellan is also credited as the father of two other concepts closely related to reader-interest classifications: the integration of reference and lending stock and the creation of separate 'popular' sections (Ainley & Totterdell 1982, p.7). According to Ainley and Totterdell (p.127), "the idea was to break away from equal funding for each category as some attract only a few readers but others attract hundreds." McClellan's formula of desired stock was used to determine the right number of books that should be provided/maintained for each category of the reader-interest classification. This formula involves variables such as the total shelf capacity of the library and the highest number of items on loan per category over a period of time (commonly the previous year). This formula was mathematically represented as follows (McClellan 1978, p.8):

$$T_s = S \left(\frac{\sqrt{L_1}}{R} \right)$$

where;

T_s = Target shelf stock for an Interest Category

S = Total circulation shelf capacity of library

L₁ = Highest number of books on loan (the Loan Component) as shown by the issue "counts" on the previous year

R = Total of all square roots for all Interest Categories 1 to n (i.e. L₁ to L_n)

On the other hand, McClellan's approach was also object of criticism. Ainley and Totterdell listed some of the disadvantages as follows: It was essentially concerned with the equitable allocation of existing resources, rather than with the ideal level of provision; unsatisfied demand was not measured -this could lead to a vicious circle, where

unsatisfied demand leads to a reduction in borrowing, which leads to a reduction in titles made available on shelves, resulting in more unsatisfied demand; other immeasurable factors -quality, condition, lighting, etc.- might also lead into the same downward spiral; and the system was devised to be operated with the Browne issue method which made tremendous demands on staff time both during the counts and for the necessary calculations which follow. Currently, this method is no more used in automated charging systems.

However, despite this criticism, McClellan's formula was used in several libraries as a method for stock control. Cheshire libraries, that adopted reader-interest classifications, were among the library systems influenced by McClellan. Here the most important motive for change was the need to assess performance of stock against demand -that is, keeping a minimum shelf stock in every identified topic of interest for a significant number of users. In 1975 most proposals for a categorized bookstock in Upton (Cheshire) were based on such factors: the arrangement of books according to majority user interests (and a better understanding of the stock/issue relationship); easier identification of dead and surplus stocks allowing transfers and control on future provision; the facility to monitor 'saturation' and 'starvation' effects within categories; easier shelving of returned books and best housekeeping; and a better understanding of the real relationship between supply and demand and a more flexible response to user needs (Astin 1982, p.14). Indeed, according to Astin, the reader interest classification and the McClellan formula in Cheshire were weighted to protect minority interests (Astin 1982, p.17). However, such a statement seems to contradict the vision that the library had of itself and the project: "Libraries with stock arranged in an order which matches the selection methods of the majority of our borrowers. Libraries guided to a uniform and high standard of presentation" (Astin 1982, p.18). In general terms, it would seem difficult to protect the interests of the minority while matching the selection methods of any majority.

Stock management thus seems to have been inextricably linked to meeting user's needs. When East Sussex County Library, for example, adopted a categorization system it declared (i) that the arrangement of stock in broad interest categories was more suited to the needs of most users than traditional classified arrangements; and (ii) that this categorization of stock should enable staff to maintain a more effective stock provision

policy (Donbroski 1982, p.79). Upon concluding the reader-interest classification project at Peacehaven and Hangleton libraries, librarians found that "deficiencies in the balance of provision could be identified through the categorised arrangement. [...] even without a fully systematic method of stock provision, reader interest arrangement of stock can, at the expense of very little effort, provide a clear guide to the relationship between levels of provision and use" (Donbroski 1982, p.83).¹³⁰

Improvements in stock control also appeared to be linked to the lack of order within categories -- an aspect commonly associated with reader-interest classifications. "Indeed librarians at both libraries have suggested that the less precise order has positive benefits in promoting greater stock knowledge amongst staff at all levels" (Donbroski 1982, p.85). In this vein, management of stock on the shelves and how this relates to user's interests was also informed by the debate over accessibility and the role of purposive and non-purposive readers found in reader-interest classifications discourses.

Also relevant -- and detailed in the literature- - was the relationship between the proper number of physical stock and a good performance in retrieval. The proper amount of stock for a good retrieval is related to the coverage (in the reader-interest classifications) of the interesting areas of stock. First, it is assumed that lower the number of stock in a library, the higher the performance in finding and retrieval by browsing. Assuming that not all the collection would be equally useful for all users, there is also the need to maintain a collection which is of interest to a particular community or user. As such, it is necessary to decide which community or users should benefit in order to keep a balanced number of stock and to improve 'retrievability.' For Astin, this is similar to the concept of "substitute borrowing" (Astin 1982, 15): "surveys have revealed that the greater majority (85%+) of public library borrowers do not require specific authors or titles but want 'good reads' in topics of related interest: i.e. they are satisfied by an adequate range of general choice in which any one title is an acceptable substitute for any other." These results were supported by a user survey carried out at Hillingdon library in the UK (Donbroski 1982, p.81), which claimed to show a high amount of substitute borrowing, and by Southwick Branch Library which found that "fewer books did not mean an immediate drop in issues" (McCarthy 1982, p.88).

Finally, Surrey Library also proposed that small and medium libraries should have very little "permanent" stock (Ainley & Totterdell 1982, p.128), since this would make more titles available (by circulating other libraries' stock) and would mean greater exposure for the "low demand" categories. The argument here seems to be that by covering categories or areas of interest for users with any kind of title, libraries are able to address the user's interests independently of the titles and the quality of those books. Depending on the philosophy of each reader-interest classification project, there might be cases where this is considered an advantage for libraries.¹³¹

Conclusions

It can be inferred from the literature that reader-interest classifications, whose focus is claimed to be on the interest of the readers, can be a method to determine the areas of the collection (i.e. the areas of the universe of knowledge to be organized with the classification system) that the users are more interested in, and therefore that must be provided with more books. While reader-interest classifications are presented as systems that collocate r distributed areas of common interest, the suggestion of using them for stock control also introduces one more variable: the degree of interest (intension) in addition to the association of subjects of common interest (extension). For example, different topics might be gathered together in a class as they might respond to a common interest of a person, but different classes (of common interest) might be of the interest of that person at different degrees. However, it is not clear what aspects of the extension of the interests, the intension of the interests, and the interests of the readers are affecting what and in which order, especially in a collection that is going to privilege the intension of the interest to determine the extension (the classes). The risk of influencing on the interests of the readers with the popularity of several topics is a shortcoming that has not been overcome by the stock control function. In this vein, it seems that aspects such as display, signaling, lighting, etc. that affect the outcome of reader-interest classifications might be behind the true factors affecting the stock control function too.

References

Ainley, P.; Totterdell, B.. 1982. *Alternative Arrangement: New Approaches to Public Library Stock*. London: Association of Assistant Librarians.

Astin, J. 1982. Cheshire: Alternative Arrangement and Beyond. In Patricia Ainley, and Barry Totterdell, eds., *Alternative Arrangement: New Approaches to Public Libraries Stock*. London: Association of Assistant Librarians, pp. 13-33.

Betts, D. 1982. Reader Interest Categories in Surrey. In Ainley, P.; Totterdell, B. eds., *Alternative Arrangement: New Approaches to Public Libraries Stock*. London: Association of Assistant Librarians, pp. 60-77.

| 132

Bragg, S. 2015. *Inventory Management: Second Edition*. Centennial, Colorado: Accounting Tools

Chandler, D. 1982. Self-service Libraries: Providing for the Smaller Community in Cambridgeshire. In: Ainley, P.; Totterdell, B. ed. *Alternative Arrangement: New Approaches to Public Libraries Stock*. London: Association of Assistant Librarians, p. 47-59.

Donbroski, L. 1982. Categorisation at East Sussex County Library. In Patricia Ainley, and Barry Totterdell, eds., *Alternative Arrangement: New Approaches to Public Libraries Stock*. London: Association of Assistant Librarians, pp. 78-85.

Martínez-Ávila, D, San Segundo, R.. 2013. Reader-Interest Classification: Concept and Terminology Historical Overview. *Knowledge organization* 40: 102-14.

Martínez-Ávila, D, San Segundo, R.; Olson, H.A.. 2014. The Use of BISAC in Libraries as New Cases of Reader-Interest Classifications. *Cataloging & classification quarterly* 52 no.2: 137-55.

McCarthy, A. 1982. Burning Issues: Stock Appeal in Sunderland. In Ainley, P., Totterdell, B. eds., *Alternative Arrangement: New Approaches to Public Libraries Stock*. London: Association of Assistant Librarians, pp. 88-100.

McClellan, A W. 1978. *The Logistics of a Public Library Bookstock*. [Harrogate]: Association of Assistant Librarians.

Miller, A. 1992. Alternative Arrangement in Glasgow City Libraries. *Public Library journal* 7 no.5: 131-3.

Rutzen, R. 1952. Shelving for Readers. *Library journal* 15: 478-82.

Sapiie, J. 1995. Reader-interest Classification: the User-friendly Schemes. *Cataloging & classification quarterly* 19 no.3/4: 143-55.

Satija, M P. 2004. *A Dictionary of Knowledge Organization*. Amritsar: Guru Nanak Dev University.

Satija, M P. 2009. Users Interest Classification. *SRELS journal of information management* 46 no.4: 353-4

Wilson, Alex. 1978. The Threshold of Choice. In *The Public Library and Its Users. Proceedings of the Public Libraries Group Weekend School, University of Warwick 25 - 28 March*. London: Library Association, Public Libraries Group.

Terminology, exhaustivity and specificity: a conceptual relation



Isadora Victorino Evangelista
São Paulo State University
isadora.biblio@marilia.unesp.br



Walter Moreira
São Paulo State University
walter.moreira@marilia.unesp.br

L33



João Batista Ernesto de Moraes
São Paulo State University
prof.joao@gmail.com

Introduction

Indexing and terminology construction are tasks that can be grouped under the common denominator of identification, description and control of the conceptual relations that the terms establish with each other. In this way, both are also concerned with conceptual analysis as a resource for the effective understanding of concepts in themselves.

Indexing, one of the processes of thematic representation, is performed with the purpose of meeting the information needs of a certain group of users. This implies that one must relate the analysis of the document itself, which is the identification and representation of its content, with the interests of the group for which it is intended. The assignment of terms that are representative of the contents of the documents and which

are, at the same time, equally representative in the specialty languages of the group of users is a terminological decision.

Indexing has two characteristics that guide it, exhaustivity and specificity. Briefly, the first relates to the extent to which one subject will be covered, while the second relates to the precision with which the terms will represent the subjects covered in the document.¹³⁴ In turn, terminology is characterized by the study of terms that are representative of a particular specialty, a knowledge domain. Terminological studies focus on the study of terms, the scope with which a domain must be researched so that the representations of their specialty languages are created.

The present article aims to verify the relationships between theoretical terminology and the concepts of exhaustivity and specificity as expressed in the Brazilian journal literature. The study is justified by the contribution of both areas to the literature and contributes so that these two universes can dialogue even more.

2 Indexing

Knowledge Organization and Representation is one of the areas of Information Science interested in the theoretical subsidies related to information treatment (BARITÉ, 2001) and also, with the procedures and systems of information representation, in order to use them effectively (HJØRLAND, 2008). To do so, some processes are needed, such as indexing.

Indexing refers to the elaboration of indexes, seeking to describe the document thematically, through terms defined from a controlled vocabulary, using representations close to the documentary content. Collison (1972) brings us two assertions related to indexing: its main purpose is to make the information contained in the informational resource available to any reader quickly, and yet, this is not a mechanical process, it is necessary that all its stages are reflected by the indexer.

Authors such as Fujita and Rubi (2003) and Boccato (2009) define indexing as one of the main processes in the library, as it is an important tool for approximating the user and the represented document. For Fujita and Rubi (2003), indexing "conditions" the results at the time of information retrieval and can, therefore, have positive or negative

effects, that is, it can either aid in locating a particular item or effectively hide it in the system.

According to ANSI/NISO Z39.19, indexing is the process of assigning terms for describing the content of documents. Also according to this standard, indexing covers all procedures of selection and organization of terms and requires human procedures,¹³⁵ although there may also be computational processes involved.

In the ABNT/NBR 12676/1992 standard, indexing is defined as the "act of identifying and describing the content of the document with terms representative of its subjects and that constitute an indexing language" (ABNT / NBR 12676/1992, 1992, p.2, our translation).

Lancaster (2004) says that there is no right way to perform indexing, but good indexing will consider beyond the issues contained in the document, why that resource is important to the user community. User communities that have different informational interests, search differently, for different terms, and therefore, the indexer must be aware that there are no single and correct set of representative terms.

Indexing has two measures to aid its effectiveness: exhaustivity and specificity, which will be addressed in the sequence.

3 Exhaustion and Specificity

The process of concept selection is intrinsic to indexing, in this case it is necessary to determine from a universe of concepts which will be the most representative of that document and one must take into account the purpose of the documentation and the users. The characteristics that best suit this purpose are exhaustivity and specificity (ABNT / NBR 12676/1992, 1992).

Exhaustivity is defined as the limits of comprehensiveness and the number of concepts representative by the terms conferred to a document. These limits of comprehensiveness should not be viewed restrictively, since a document used by one group of users may also be used by another group with different interests (ABNT / NBR 12676/1992, 1992; LANCASTER, 2004).

For Maturana (1992), exhaustivity is the extension ability with which the different themes of a document will be approached, identified from the documentary analysis by the indexer and, later, translated from a documentary language when it then becomes, representative terms.

As for specificity, we can say that it refers to the level of precision the terms will have; how specifically the subjects will be covered. For Lancaster (2004), exhaustivity takes into consideration the approached thematic extension, specificity considers the thematic depth of representation, while the first measure may be considered an indexing policy decision, the second regards the documentary language.

According to NISO TR02: 1997 standard (1997), specificity is the proximity between the defined term and the document subject. This standard gives an example: "pick-up style trucks" can be used for that particular type of truck while the term "truck" would be representative of all types of trucks. Still according to the standard, exhaustivity and specificity can be combined to determine the depth of the representation.

We can say that these two processes are complementary in indexing, so that the thematic content of the document is completely covered and that the terms that will represent it, are accurate. In the next section, some theoretical assumptions about terminological studies are presented and later, an analysis of how the concepts of exhaustivity and specificity relate to this universe.

4 Terminological Studies

Terminological studies begin with Eugen Wuster in the 1930s, considered the founder of modern terminology and one of the leading representatives of Vienna school. Wuster defines terminology before three assertions: an arrangement of concepts proper to a specialized domain and its denominations, the specialized lexicology of a domain and *the general theory of terminology*, when he affirms that in terminology the principles are common even in varied domains, regardless the language (FAULSTICH, 2001).

Another relevant researcher in this area is Maria Tereza Cabré, who defines terminology as being "the discipline whose object is the study and compilation of specialized terms" (CABRÉ, 1993, p.21). For the author, although the study of

terminology, as we know it today, is relatively recent, beginning in the mid-1930s, it is enough to recall Lavoisier's works in chemistry and Lineu's in Botany, to see that there is interest in denominations of scientific concepts by the experts.

For the author, the concern of linguists in recent years is focused on constructing a theory that deals with the principles of real and possible languages and does not pay attention to the communicational aspects of instruments. In a communicative conception of language, terminology could aid the work of linguists.

In this sense, Cabré develops, together with the group of researchers of Pompeu Fabra University, Spain, the Communicative Theory of Terminology, in which fundamentals are presented and structured from the communicative focus of specialized languages, with a greater concern than the normalized purposes (KRIEGER, 2000).

In the 1990s, studies on terminologies were intensified, considering the complexity involved in their functioning, receiving contributions from areas such as sociolinguistics, translatology and even from the field of artificial intelligence (KRIEGER, 2000).

In a study carried out by Cervantes, Fujita and Rubi (2008), the authors consider the study of terminological sets in the areas of science to be of great importance, and this concern with specialized languages is important as, in addition to determining the domain, they are able to assist in the knowledge transmission.

Krieger (2000) points out that terminologies, as linguistic systems, are not neutral, since they are derived from natural languages and express specialized knowledge, and in addition, are not absolutely artificial, as they seek to construct themselves from language, and later identify precise concepts in different fields of knowledge, from terms, avoiding polysemy and ambiguities.

5 Methodological procedures

The proposed research has an exploratory character, as it is presented as a preliminary study so that the idea of the objective is presented and the hypotheses are formulated later. It is also classified as qualitative, considering that the object of research were articles that prove the relation of terminology to the concepts of exhaustivity and specificity.

To establish the corpus of research, Scielo database was used and the following terms were used, in Portuguese language, for the search: "terminology", " exhaustivity " and "specificity". These terms were searched in combination, through advanced search, in the areas of higher information content: title, abstract and keywords, and there was no temporal delimitation. The journals *Ciência da Informação*, *Perspectivas em Ciência da Informação* e *Transinformação* were pre-defined, because they are recognized journals in the area and would bring the terms within the universe of research of Information Science. With these search terms, no journals treating exclusively with indexing or terminology were retrieved. The corpus totaled sixteen articles, of which seven had concepts of terminology that could be used in the research.

Subsequently, a table with the definitions of the three terms were elaborated, in a comparative way.

6 Results

Below, the comparative table is presented, demonstrating that concepts of terminology can correspond to the concepts of exhaustivity and specificity.

It can be noticed that the two areas - indexing and terminology - present convergence between the concepts, identifying that a greater dialogue between them is possible. Considering that both fields are concerned with the elaboration of representative terms, having therefore the same objective, characteristics such as exhaustivity and specificity must be present at the time of this elaboration.

Table 2 - Comparison between the concepts of terminology, exhaustivity and specificity.

TERMINOLOGY	EXHAUSTIVITY	SPECIFICITY
“In terminology, emphasis is placed on the development of glossaries, which are instruments for standardizing terms of a particular specialty or domain.” (SALES; CAMPOS; GOMES, 2008, p. 66)	Regarding a domain, it must then be exhaustive.	Terms referring to a specialty should be as specific as the given specialty.
“(Terminology is) any activity related to the systematization	Exhaustivity refers to the number of concepts that will	

and representation of concepts or presentation of terms based on established principles and methods as well as a set of terms that constitute a system of concepts of a given area” (DIAS, 2000, p. 90)	be established to the terms, constituting, therefore, a system of concepts.	
“(…) as a product, is a set of terms, or vocabulary, of a particular specialty.” (DIAS, 2000, p. 90)	Exhaustivity refers to the number of terms that will be representative of a given document.	Because it is a specialty, it is therefore specific.
“If taken as an object, that is, terminology as a set of terms of a specialty, it is noted that each area of knowledge addresses its terms differently.” (DIAS, 2000, p. 90)		In the case of a specialty, the terms must then be specific.
“A specialized terminology gives indexer subsidies to more accurately represent documents that will be available in databases for the scientific community conduct their searches.” (MARTINS; CARVALHO, 2014, p. 122)	Exhaustivity is, in the same way, an activity that subsidizes the work of the indexer.	Specificity is, in the same way, an activity that subsidizes the work of the indexer, in addition to being directly related to the accuracy of the representation.
“Terminology, as a means of formal communication among specialists of a domain, is constituted by monoreferential terms: each term has a specific concept and is related to the others by its definition, allowing the precision and normalization of this language.” (ALCAIDE et al., 2001, p. 26)	Because it represents a domain, it must therefore be exhaustive.	Each term having a specific term requires specificity. Moreover, one of the principles of specificity is precisely the accuracy of the retrieval.
“The term is a unit of communication that represents the concept; its meaning is denotative, composed by words determined by experts from a specific area of knowledge.” (BORBA; VAN DER LAAN; CHINI, 2012, p. 32)		Thus, the term representative of a specific area of knowledge requires specificity.
“(…) knowledge of the subject (that is, of the concepts of a certain area)	To know the subject and the concepts of a certain area, they must thus be exhaustive.	

is guiding element for the creation of the term.” (BORBA; VAN DER LAAN; CHINI, 2012, p. 32)		
“Terminology employed in scientific evaluation can be framed among qualitative indicators, because among its strengths is its relation with specific contents of the scientific ideas represented by the terms.” (MORAES, 2007, p. 33)		Because it carries specific content, it requires specificity.
“The General Theory of Terminology seeks objectivity and precision, based on analytical philosophy.” (MORAES, 2007, p. 34)		Specificity is directly linked to the precision of the terms.
“To remedy this problem, a terminological operational standardization was initiated without, however, failing to respect the specificity of the term” (MORAES, 2007, p. 35)		Again, the author refers to the notion of specificity.
“There were cases of terms that (...) appeared with too much specificity. However, Fiocruz researchers opted for a more comprehensive term.” (MORAES, 2007, p. 36)	In certain knowledge domains, at times the terms should be broader and therefore more exhaustive.	
“However, even when it is not a question of specialized universes, that is, they are universes whose boundaries are not well defined, the terminological principles associated with the documentary ones are essential to refer to the process of spatial and visual organization of the repertoire at stake” (LARA, 2004, p. 243)	Again, some of knowledge domains are not delimited, and therefore a more comprehensive and exhaustive terminology is needed.	

Source: by the authors

7 Conclusion

The present work sought to demonstrate how the concepts of terminology can

correlate with the concepts of exhaustivity and specificity found in indexing. This search was based on the Brazilian literature in the area of Information Science and resulted in a comparative table, in which excerpts from the papers were extracted, when these correlations were found.

It was noted that some concepts of terminology can, therefore, be related to the¹⁴¹ basic concepts of exhaustivity and specificity in indexing. Especially considering that a terminology must be exhaustive to understand an entire knowledge domain and also specific to well represent the specialties of a given knowledge.

This exploratory research sought only to demonstrate that there are evidences of correlation and, therefore, further research is suggested in order to seek these concepts in the papers by the main theoreticians of the area.

Acknowledgement: Fundação de Amparo à Pesquisa do Estado de São Paulo.

References

The references was made following the ABNT rules.

ALCAIDE, G. S. et al. Análise comparativa e de consistência entre representação automática e manual de informações documentárias. *Transinformação*, v. 13, n. 1, p. 23-41, 2006.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. *NBR 12676: Métodos para análise de documentos - Determinação de seus assuntos e seleção de termos de indexação - Procedimento*. Rio de Janeiro: ABNT, 1992.

BARITÉ, M. Organización del conocimiento: un nuevo marco teórico-conceptual en Bibliotecología y Documentación. In: CARRARA, K. (org.). *Educação, Universidade e Pesquisa*. Marília: Unesp-Marília-Publicações; São Paulo: FAPESP, 2001. p.35-60.

BORBA, D. S.; VAN DER LAAN, R. H.; CHINI, B. R. Palavras-chave: convergências e diferenciações entre a linguagem natural e a terminologia. *Perspectivas em Ciência da Informação*, v. 17, n. 2, p. 26-36, 2012.

CABRÉ, M. T. *La terminología: teoría, metodología, aplicaciones*. Barcelona: Antártida/Empúries, 1993. 529p

CERVANTES, B. M. N.; RUBI, M. P.; FUJITA, M. S. L. Terminologias em política de indexação. *Ibersid (Zaragoza)*, v. 2, p. 211-222, 2008.

COLLISON, R. L. *Índices e indexação: guia para indexação de livros e coleções de livros, periódicos, partituras musicais, discos, filmes e outros materiais*. Tradução por Antônio Agenor Birquet de Lemos. São Paulo: Polígono, 1972, 223 p.

DIAS, C. A. Terminologia: conceitos e aplicações. *Ciência da Informação*, v. 29, n. 1, p. 90-92, 2000.

FAULSTICH, E. Aspectos de terminologia geral e terminologia variacionista. *TRADTERM*, v. 7, São Paulo, 2001.

| 142

FUJITA, M. S. L.; RUBI, M. P. Elementos de política de indexação em manuais de indexação de sistemas de informação especializados. *Perspectivas em Ciência da Informação*, v. 8, n.8, p. 38-54, 2003.

FUJITA, M. S. L. (Org). *A indexação de livros a partir da percepção de catalogadores e usuários de bibliotecas universitárias: um estudo de observação do contexto sociocognitivo com protocolos verbais*. São Paulo: Cultura Acadêmica, 2009.

HJØRLAND, B. What is knowledge organization?. *Knowledge Organization*, v. 35, n. 3/2, 86-101, 2008.

KRIEGER, M. G. Terminologia revisitada. *D.E.L.T.A.*, v. 16, n. 2, 209-228, 2000.

LANCASTER, F. W. *Indexação e resumos: teoria e prática*. Tradução de Antonio Agenor Briquet de Lemos. Brasília: Briquet de Lemos, 2004.

LARA, M. L. G. Linguagem documentária e terminologia. *Transinformação*, v. 16, n. 3, p. 231-240, 2004.

MARTINS, E. C. D.; CARVALHO, T. Recuperação da informação em psicologia: LILACS e Index Psi Revistas Tecno-Científicas. *Perspectivas em Ciência da Informação*, v. 19, n. 2, p. 118-130, 2014.

MATURANA, M. T. I. *Analysis y recuperación de información documental*. Santiago: Ediciones Instituto Profesional de Santiago, 1992, 237 p.

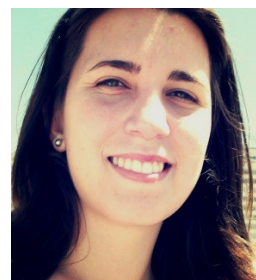
MORAES, A. F. Terminologia como indicador qualitativo. *Transinformação*, v. 19, n. 1, p. 31-38, 2007.

NATIONAL INFORMATION STANDARDS ORGANIZATION. *ANSI/NISO Z39.19: guidelines for the construction, format, and management of monolingual controlled vocabularies*. Estados Unidos: NISO, 2010.

NATIONAL INFORMATION STANDARDS ORGANIZATION. *NISO TR02: guidelines for indexes and related information retrieval devices*. Estados Unidos: NISO, 1997.

SALES, L. F.; CAMPOS, M. L. A.; GOMES, H. E. Ontologias de domínio: um estudo das relações conceituais. *Perspectivas em Ciência da Informação*, v. 13, n. 2, p. 62-76, 2008.

Aspects of subject metadata in cataloging photographs



43

Ana Carolina Simionato
Federal University of São Carlos
acsimionato@ufscar.br

Introduction

The emergence of new possibilities for the localization of images in digital media revealed that some forms of representation still remain with flaws that hinder their search and access. Whether in the traditional or digital environment, this fact lacks discussions, due to the great difficulties in defining metadata, descriptive and subject for photographs.

In their particularities, photographs must be considered as a primordial point, that is, the informational resource must be the front of any form of representation. Preferably not being tied to a general rule that concentrates diverse types of informational resources, such as general rules, but rather, conceptualizing the solution of the image type.

This paper aims to verify the aspects of subject analysis in the representation of photographs, in order to present forms and instruments that assist in the definition of the value for the field, subject metadata, including the characteristics of the knowledge and information organization systems.

In this sense, this paper does not aim to discuss the theoretical perspectives of development of each instrument for photograph representation, but to point them as possibilities an efficient way for localization, retrieval, access and use. Thus, the methodology used is theoretical, exploratory and qualitative in relation to the dynamics of the world and the subject, due to the interpretation and understanding of the phenomena. (GIL, 2002).

This paper corresponds to a completed part of the ongoing research on the thematic and descriptive treatment of photographic collections of the state of Paraná - Brazil, developed by the research group Information and Knowledge Organization and

Representation of Image Resources - ORICRI. Through the bibliographical research carried out, we found several ways of accomplishing the attribution of value of the subject metadata, however, no joint guidelines were found to be performed systematically to what is found in Information Science or even, unknown to institutions representing photographs.

The aspects of subject metadata in photography cataloging have been divided into¹⁴⁴ topics that designate photograph representation, the possibilities for subject metadata and the conclusions of the study.

2 Photograph Representation

From the invention of the dark camera in the nineteenth century and the guarantee of fixation on a plate for commercial processes in 1839, known as daguerreotype, the photographic record was popularized with portable cameras of reduced cost that today have been transposed to the digital medium.

Consequently, with the increase in the number of cameras, the proportion of photographic records was exponential, took proportions in which they enabled the recognition of their historical and cultural value, justifying the storage in information units. Through storing this new collection possibility, difficulties arise in the representation of these informational resources.

Digital cameras and their images deal fluently with metadata, attributes that characterize an informational resource, in this case, image resources. It is possible to say that metadata are the simplest forms of description in digital informational environments, they are composed of attributes and values, previously defined by a methodological form for composition of a metadata pattern.

Digital images automatically possess metadata that describe them in a technical way, named Exchangeable Image File Format (EXIF) data, these metadata are information recorded together with the record in relation to 3.5 mm of focal length, for example. However, in traditional photographs, professionals must choose the best way to describe the technical information and in this case, regardless of the medium, both should elucidate the representation by subject.

Representation can be conceptualized as a set of syntactic and semantic conventions that make the description of things possible in an intrinsic and extrinsic way, and it requires patterns to formalize and adapt the presented and proposed presentation to the user in order to improve his searches, access and location. In this sense, Smit (1996, p. 29) points out that

The proposition of a methodology for the analysis of photography presupposes an understanding of the essence of photography, of what characterizes it, of the reasons why it is produced and, above all, of the conditions in which it will be used. In other words, it becomes necessary to understand the photographic image as information to be treated and retrieved.

However, the standardization sought by librarians is not the guarantee of adequate representation for a photograph and the representation instruments still do not offer efficient support. (SIMIONATO, 2012).

The traditional codes of Library science cannot cover the specificity of image. For example, the Anglo-American Cataloging Code (AACR2r), which does not present necessary subsidies for the description of photographic techniques and main subject. With this, the professional should always consult other metadata standards, such as CDWA, VRA Core, Cataloging Cultural Objects, Graphic Materials or even the institution's decisive catalog.

If the institution uses MARC 21, the available fields and subfields will find subject metadata in group 6XX, corresponding to 650 for topic term and 690 for the term assigned to the institution.

In this context, what opens the discussion and the theme of this paper is that the structures currently established correspond in standards and methodologies accepted and facilitated to librarians, however, information corresponding to the value of metadata, which, in many cases, makes it difficult to understand, and without the proper standardization of instructions for completion, affect interoperability.

3 The possibilities for Subject Metadata

The representation of an image for photographers is named metalogging, which consists of the inclusion of information in a way that can be stored and used as an aid for retrieving that image from a database or collection after being registered. (DALY, 2014). Since much of the difficulty in filling metadata is in the subject field, which photographers call the keywording process, assigning terms that increase visibility and allow more relevant searches is necessary. (DALY, 2014).¹⁴⁶

This field depends on the possibilities that the environment or even the institution offers, for this reason, many forms of representation by subject are totally free to the user, so called folksonomy.

Among some digital environments such as Flickr, Picasa, Zoomr, Pixlr, among others, folksonomy was chosen as the main form of retrieval attributing to users the responsibility of terms, through natural language terms. In this way, users categorize and index themes that they judge to be related to an image, however, most times these tags end up representing content that is not what was presented. As Herst et al. (2007, unpagged, our translation) point out that "[...] a computer searches for images on the Internet through keywords or tags, but not always finds what is desired; this happens, as images are found based on the metadata that accompany them [...]".

In thinking that folksonomy and keywording in many cases fail to ensure the best image retrieval, information professionals present some of the tools that can be part of this process of subject definition.

It is worth pointing out that the need to create attributes or even the appropriate assignment of representative values to this information, so that they are adequately presented to the users, through an efficient and effective retrieval. That is, the instruments for standardizing this information avoid some misunderstandings such as using 'white and black', a descriptive value in a thematic value.

For this, it is necessary to make an analysis and synthesis of the image content and later to carry out the description and the metadata, these procedures are inserted the main operations of the documentary analysis, being also included the processes of classification, indexation and documentary condensation.

According to the authors Boccato and Fujita (2006, p.90)

Documentary analysis of images must comply with the precepts of documentation, reflecting the credibility and security at the moment of information retrieval by the user. Thus, keeping with the central objective of documentary analysis, documentary information should promote the identification of informational materials that satisfactorily respond to users' questions and, on the other hand, make it possible to make decisions about the consultation and choice of a given original document. | 147

Documentary analysis should be based mainly on an analysis of the concrete connotative (RODRIGUES, 2011) for interpretation of the photograph must be observed by its constituent elements, with terms that are clear, for example: architecture, city hall, city.

This analysis, using the concrete connotative, reduces the subjectivity of the analysis and description of the content. Erwin Panofsky (1979) established levels for analyzing the image that guides this subjectivity. Panofsky (1979) presents three levels for image analysis: 1) pre-iconographic, where the objects and actions represented by the image are described generally; 2) iconographic, establishes the secondary or conventional subject illustrated by the image. That is, it is the determination of the abstract or symbolic meaning of the image, synthesized from its component elements, detected by pre-iconographic analysis; and 3) iconological, which proposes an interpretation of the intrinsic meaning of the content of the image. Iconological analysis is constructed from previous levels, but it has influences from the analyst's knowledge about the cultural, artistic and social environment in which the image was generated.

The image is both specific and generic, according to Shatford (1986) based on the levels of Panofsky (1979) discusses image representation, introducing a distinction between generic, specific. Shatford (1986) represents levels in categories, as a user can understand the pre-iconographic sense of an image (e.g. river), but he cannot understand the needs in iconographic terms (e.g. Amazon River).

The categories by Shatford (1986) represent the questions of: Who? (Beings), Where? (Space), When? (Time), How? (Technique) and What? (action). These categories were also applied and studied by other authors, such as Smit (1989), Kossoy (2001), Manini (2001) and Maimone and Gracioso (2007).

In the study by Costa (2008), these categories were also related to categories Personality, Matter, Energy, Space and Time - PMEST, called 'essential categories',

proposed by Ranganathan in 1931. Thus, finding relation with Personality - Who? Matter - What?, Energy - How?, Space - Where? And finally, Time - When ?.

Therefore, documentary analysis for assigning value to subject metadata, there are other linked processes that involve beyond categories, the process is indexing. From indexing, a process originating from the need for a deepening of specific contents, it is possible to identify the theme of a document and represent it through concepts, that is, transform the natural language - LN into a documentary language - LD. (FUJITA, 2009).¹⁴⁸

From this terminological process, instruments for standardization of terms are created, these instruments vary in their level of deepening to the content as well as their form of terminological standardization.

Among the instruments, controlled vocabularies stand out because they are knowledge organization systems (BOCCATO; RUBI; FUJITA, 2014), and that allow the representation of concepts identified and selected through the concrete connotative analysis of the photographic content.

According to Harpring (2010, p. 166, our translation) the controlled vocabulary in "[...] cataloging is a more efficient way of incorporating terminologies, through the use of local authorities or including controlled vocabularies published in their entirety .".

Thus, controlled vocabularies are arranged in a specific order and are characterized by three relationships: equivalence, hierarchical and associative. When their terms are structured in a hierarchical way, they are consistent with the formation of a thesaurus, which can be both monolingual and multilingual, and the objectives focus on promoting consistency in content indexing and facilitating search and navigation. (HARPRING, 2010).

Controlled vocabularies for image resources have terminological structure and repertoire that enable representation and analysis for insertion into subject metadata, in which case, will be exemplified by the initiative of the Getty Research Institute in the United States.

Getty Research Institute built the set of the Art & Architecture Thesaurus (AAT), the Getty Thesaurus of Geographic Names (TGN) and under construction The Cultural Objects Name Authority (ULAN), which are all structured vocabularies that can be used to improve access to Information on art, architecture, and material culture. The thesaurus

terms and controlled vocabularies of the Getty Research Institute are collected or are related from collections of museums, special collections, archives, libraries, academic research, and other sources. Thesauri can be accessed through the link: <<http://www.getty.edu/research/tools/vocabularies/>>.

Another important feature is that they are being built to allow their use in linked data,¹⁴⁹ working together with the Linked Open Data initiative, making part of the Getty Research Institute's ongoing effort to make knowledge resources freely available to everyone.

4 Conclusion

Among the notes already made during the paper, they elucidate that documentary analysis, controlled vocabularies and thesauri increase metadata consistency in retrieval systems for the location, access, use and reuse of photographs, in databases, catalogs and other forms of image storage.

In this sense, it is important for information professionals to seriously discuss the need to use more structured ways to fill in subject metadata, as well as discuss with their main users, photographers and other specialized users, a transformation of terms that are naturally used so that the system is better and reflects their interlocution.

Therefore, the use of terminological structures, hierarchical or not, provide important tools for the retrieval of image resources, be it in different places, in different languages and this may lead to new and increasingly technological proposals for the linking and relationship of registers, in line with the Linked Data proposal.

References

The references was made following the ABNT rules.

BOCCATO, V. R. C.; FUJITA, M. S. L. Discutindo a análise documental de fotografias: uma síntese bibliográfica. *Cadernos BAD*, n. 2, 2006, p. 84-100.

BOCCATO, V. R. C.; RUBI, M. P.; FUJITA, M. S. L. Aplicabilidade de vocabulário controlado na análise documental de fotografias: uma prática possível. In: SOUZA, L. M. A.; FUJITA, M. S. L.; GRACIOSO, L. (Org.). *A imagem em Ciência da Informação: reflexões teóricas e experiências práticas*. Marília: Oficina Universitária; Cultura Acadêmica, 2014, p. 117-152.

COSTA, L. S. F. *Uma contribuição da teoria literária para a análise de conteúdo de imagens publicitárias do fim do século XIX e primeira metade do século XX, contemplando aspectos da natureza brasileira*. 263f. 2008. Tese (Doutorado em Ciência da Informação) - Universidade Estadual Paulista "Júlio de Mesquita Filho", Faculdade de Filosofia e Ciências, 2008.

DALY, T. *The fundamentals of digital photography*. Londres: A&C Black, 2014.

FUJITA, M. S. L. (Org.) *A indexação de livros: a percepção de catalogadores e usuários de bibliotecas universitárias*. São Paulo: Cultura Acadêmica, 2009. Available at: http://www.culturaacademica.com.br/catalogo-detalle.asp?ctl_id=56. Access at: 20 mai. 2015.

GIL, A. C. *Metodologia da pesquisa*. São Paulo: Atlas, 2002.

HARPRING, P. *Introduction to controlled vocabularies: terminology for art, architecture, and other cultural works*. Los Angeles: Getty Publications, 2010.

HEARST, M. et al. *Faceted metadata for image search and browsing*. Berkeley: University of California, 2007.

KOSSOY, B. *Fotografia & história*. São Paulo: Ateliê, 2001.

MAIMONE, G. D.; GRACIOSO, L. S. Representação temática de imagens: perspectivas metodológicas. *Informação & Informação*, v. 12, n. 1, 2007. Available at: <http://www.uel.br/revistas/uel/index.php/informacao/article/viewFile/1760/1504>. Access at: 20 mai. 2015.

MANINI, M. P. Análise documentária de imagens. *Informação & Sociedade*, v. 11, n. 1, 2001. Available at: <http://www.ies.ufpb.br/ojs2/index.php/ies/article/view/313/236>. Access at: 20 mai. 2015.

PANOFSKY, E. *Significado nas artes visuais*. 2.ed. São Paulo: Perspectiva, 1979.

RODRIGUES, R. C. *Análise e tematização da imagem fotográfica: determinação, delimitação e direcionamento dos discursos da imagem fotográfica*. 2011. Tese (Doutorado em Ciência da Informação) – Universidade de Brasília, 2011.

SHATFORD, S. Analyzing the subject of a picture: a theoretical approach. *Cataloging & Classification*, v. 6, n. 3, p. 39-62, 1986.

SIMIONATO, A. C. *Representação, acesso, uso e reuso da imagem digital*. 142f. 2012. Dissertação (Mestrado em Ciência da Informação) - Universidade Estadual Paulista "Júlio de Mesquita Filho", Faculdade de Filosofia e Ciências, Marília/SP, 2012.

SMIT, J. W. A análise da imagem: um primeiro plano. In: SMIT, J. W. (Org.) *Análise documentária: a análise da síntese*. Brasília: IBICT, 1989.

SMIT, J. W. A representação da imagem. *Informare: cadernos do Programa de Pós-graduação em Ciência da Informação*, v.2, p. 28 -36, 1996.

Documentary languages and audiovisual resources treatment: theoretical and methodological interfaces



Francisnaira Cristina Ravazi
Regional Direction of Education of Tupa
francisnaira@gmail.com



Walter Moreira
São Paulo State University
walter.moreira@marilia.unesp.br

L51

Introduction

Throughout history, librarianship and documentation have been concerned with the development of instruments for organizing and representing information recorded in printed documents. Audiovisual resources received less attention for various reasons. In this sense, there are no instruments of classification, for example, that consider the specificities of a scenario that involves "electronic audiovisual culture". Among other things, the offer of other ways of reading and perceiving question the centrality of the school-book relationship as the nucleus of the current school system (MARTÍN-BARBERO; REY, 1999).

It is possible to infer that this (relatively) new model of culture gives the user information, especially the younger ones, new ways of reading that naturally interfere with their worldview and how they will consume information at its different levels.

The advances made in theories and methodologies developed by librarianship and documentation for knowledge organization, considering the perspective of their social use, are limited in relation to the specificities of audiovisual language. In the fields of communication and pedagogy, there are some considerable theoretical advances, although there is, in this case, concerns with what can be called professional organization of audiovisual resources.

Given the importance and the popularity of audiovisual resources, it is natural that there are anticipations of solutions, by the most diverse sectors interested in the question,

in relation to the more classic areas concerned with the information organization, representation and retrieval. It is necessary and extremely useful to information science, however, to identify, describe and analyze the models used - professionally or not - in documentary representation of audiovisual resources.

We propose, as a general objective of this paper, to identify the area of theoretical and methodological confluence between documentary languages and the treatment of audiovisual resources. The general objective is complemented operationally with the following specific objectives: a) to discuss the audiovisual concept and b) to identify and compare, through systematization, the methodologies for organization of audiovisual information.

Two types of corpora were used to carry out the research: a) documentary corpus: from which the articles and studies for analysis were extracted; b) corpus of analysis: the articles and selected papers from the documentary corpus through search strategies that met the objectives of this research.

The documentary corpus was composed by the whole set of journals available in the Reference Database of Journal Articles in Information Science (BRAPCI) and in the proceedings of ENANCIB. In both cases, a 13-year chronological cut was observed during the period 2000 - 2012.

In the search for composing the analysis corpus, we used strategies that included the following terms: "classification", "indexing language", "documentary language", "document language", "information organization", "audiovisual", "audiovisual resource", "knowledge organization", "information representation", "knowledge representation", "thesaurus", "ontology". Preference was given to the presence of the terms in the fields of keywords and titles of the papers and, in a complementary way, in the abstracts. The terms were proposed in isolation and with qualifiers, considering also their variations of number.

For the analysis, the localized papers were categorized by theme: a) documentary languages in general; b) classification systems; c) thesauri; d) ontologies; e) audiovisual. As research interests lies in particular on the papers contemplating theories and/or methodologies on the organization of audiovisual information, four more categories were formed due to the intertwining of the previous ones: f) documentary and audiovisual

languages; g) classification and audiovisual systems; c) thesauri and audiovisual; D) ontologies and audiovisual.

After categorizing the articles, the analysis was developed with the purpose of identifying concepts, theoretical contributions and methodologies applied in the solution of problems related to the organization and representation of audiovisual information. | 153

Audiovisual concept

The concept of audiovisual is still not stable and its identification required a careful analysis of the literature, given its centrality in this research.

Using the search expression "audiovisual" in BRAPCI, 27 articles were identified in the study theme. In using other proposed search terms ("audio visual", "audiovisual" and "audio visuals"), fewer articles were found, but the same as those already been retrieved with the term "audiovisual", proving the terminological oscillation that exists on the theme as observed by several authors, among them, Bethônico (2006). This oscillation in the availability of information hinders users when information is retrieved and obviously makes bibliographic research difficult, as well as compromising the accuracy of the concept.

The Dictionary of Librarianship and Archives defines audiovisual material as an "information medium that cannot do without hearing or vision equipment that reveals its content" (CUNHA; CAVALCANTI, 2008, p 241). The Brazilian Dictionary of Archival Terminology classifies audiovisual documentation as "a documentary genre composed of documents that contain images, still or moving, and sound records, such as films and videomagnetic tapes" (ARQUIVO NACIONAL, 2005, p.73). Also in the Dictionary of Librarianship and Archives, the term audiovisual document is defined as "a document that reproduces still or moving images, as well as sound records on any medium, and which requires appropriate equipment to be visualized or executed" (CUNHA, CAVALCANTI, 2008, P 133). In both dictionaries, audiovisual document is described as a document that presents both sound and image at the same time, therefore different from what is described in the term audiovisual material.

The International Encyclopedia, edited by Feather and Sturges (2003, free translation), defines audiovisual materials as a generic term to describe the content of

information made available in storage and transmission media and formats that use images and sound instead of, or sometimes in addition to, textual matter. It includes: audio CD, discs and tapes, photos, slides, films and videos and formats that can combine two or more of these [...]. The term, however, is still quite useful, and has not been entirely replaced by multimedia, which is often misused as a fashion synonymous. | 154

Some authors (CEBRIAN-HERREROS, 1983; BETHÔNICO, 2006) argue that audiovisuals must simultaneously present image and sound. In this line, Bethônico (2006, p.59) argues that silent cinema, for example, "is not really audiovisual without the music that participates in the construction of climates and that was formerly played live in the sessions."

Other scholars question the simultaneity of image and sound, such as Araújo (1992), Smit (1993) and Dieuzeide (1965). The latter defines audiovisuals as "mechanical or electronic means of recording, reproducing and disseminating sound or visual messages used separately or jointly to present knowledge, facilitate its acquisition or modify certain behaviors."

However, in what characterizes audiovisual, the question of presenting sound and image at the same time is treated by almost all the authors who work with this resource, without reaching a consensus on the issue.

Regarding the scope of the concept, it is shared in this work, from the conception that audiovisual needs to present sound and image at the same time. Considered audiovisual, therefore, resources such as movies, videos, video clips and others, since they present as characteristic the simultaneous presence of sound and image.

Bibliographic base of audiovisual resources

For the analysis of the bibliographic base, all the texts retrieved in the bibliographic survey, in BRAPCI and ENANCIB, were selected using the search term "audiovisual". In each article the bibliographic references were analyzed in order to identify the area of knowledge in which they are located.

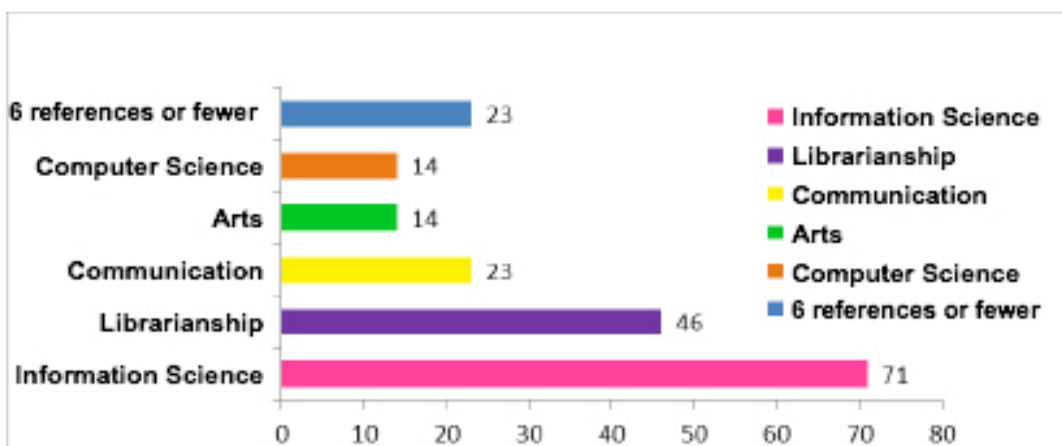
The authors/works were grouped and ordered by the criterion of the number of times they were referenced. In this way, two large groups were identified. The first one,

with 18 authors, whose works were referenced twice. The second group with 173 authors/works that received only one reference.

Considering that the analyzed texts addressed the theme audiovisual, it is expressive that the maximum number of citations one same text has received is equal to 2, it is equally expressive the number of authors who received only one citation, that is, they were mentioned only once. In this case, it is possible to infer that there is no consensus on the literature that bases the research on audiovisual.

The references used in the selected papers were also classified in areas, according to the table of areas from the National Council of Scientific and Technological Development (CNPq), with the objective of mapping the areas of foundation for studies on audiovisual resources. The result is expressed in Figure 1. The areas that received 6 or fewer references were grouped in the same item, in order to improve the visualization.

Figure 1 - Classification of references used in texts about audiovisual



Source: Research Data

Information Science and Librarianship have offered the largest subsidies for audiovisual studies. This can be taken as natural in this study, since the works published in journals and scientific events of these areas were used as data. On the other hand, it should also be noted that this may be a sign of lack of dialogue. The issue of audiovisuals has been widely discussed in Communication and also in Education.

Instruments used in the organization of audiovisual resources

The results presented in the sequence refer to the systematization of the larger set of data used in previous research work (RAVAZI, 2014).

Chart 1 presents the typology of information organization tools used for the information treatment of audiovisual resources in the analyzed texts. The bibliographic references used in the first column of the chart were abbreviated for visualization matters and appear completely in the list of references at the end of the paper.

Chart 1 - Typology of used instruments

REFERENCE	INSTRUMENT
FAGUNDES; PRADO (1999). Videoteca digital. [Digital videolibrary]	Data base made with <i>Microsoft Office</i> resources
GONÇALVES (1992). Os novos paradigmas da imagem em movimento. [The new paradigms of moving image]	Thesaurus
MOURA et al. (2005). Linguagens de indexação em contextos cinematográficos. [Indexing languages in cinematographic contexts.]	Thesaurus
BARRETO (2007). Desafios e avanços na recuperação automática da informação audiovisual. [Challenges and advances in automatic recovery of audiovisual information.]	Metadata
MIRANDA; GARBELINI (2011). Tratamento técnico da documentação audiovisual na TV da Universidade Federal de Goiás. [Technical treatment of audiovisual documentation on TV of Federal University of Goiás.]	Collection analysis, , decoupage, elaboration of tags and preservation.
SOUSA; DUQUE (2012). Uso de ontologia para recuperação da informação disponibilizada em vídeos por meio de indexação multimodal. [Use of ontology to retrieve information made available in videos through multimodal indexing.]	Multimodal indexing and ontology
MUCHERONI; RIBEIRO; PAIVA (2012). Descrição dinâmica para documentos multimodais em ambiente digital. [Dynamic description for multimodal documents in digital environment.]	Ontology for multimodal documents
SILVA; SOUZA (2012). Vocabulários e metadados para descrição de recursos	Vocabulary (including ontologies) and metadata based on <i>linked data</i>

multimídia. [Vocabularies and metadata for multimedia resource description.]	
ALVARENGA; SOUZA (2012). A Universidade Federal de Minas Gerais no contexto do acesso aberto à informação científica. [The Federal University of Minas Gerais in the context of open access to scientific information.]	Tags
SOUSA; DUQUE (2011). Recuperação da informação disponibilizada em vídeos: uma abordagem utilizando ontologia, processamento de linguagem natural e descritores multimodais. [Information retrieval made available in videos: an approach using ontology, natural language processing and multimodal descriptors.]	Ontology and multimodal indexing

Source: Research Data

As observed from Chart 1, some authors use relatively traditional library tools, such as thesauri and metadata, while others use ontologies and multimodal indexing, both also relatively new objects of study in the field of information science. It is also observed that instruments or techniques that do not use librarianship tools, such as is the case of the studies by Fagundes and Prado (1999) and Miranda and Garbelini (2011).

It may be inferred that authors who have used instruments other than those traditionally assigned to information science are not researchers of information science themes or who do not use its theoretical framework. It was observed, however, that all the authors mentioned in Chart 1 have a direct and close relationship with information science, as was verified in the field of affiliation and field of action of each studied text.

Conclusion

Through the analysis carried out in accordance with the objectives of the paper, it was possible to verify the concept and the bibliographic basis that support the conceptions of organization of audiovisual information. The procedures for the organization of this type of resource are carried out both through information science tools (thesauri, metadata, multimodal indexing, ontology and tags) and other means (database using text editor, decoupage, tag elaboration).

It has been found that the concept of audiovisual is still not sufficiently clear. There is still no precise answer to the identification of the concept and not even to a typology that precisely describes what these materials are and what their specificities are. It was observed that even at the syntactic level, the term is still not absolutely stable, as it was possible to verify by the presence of varied forms of spelling in journal articles of an area that, it is worth remembering, excels by information organization and vocabulary control. ¹⁵⁸

It is necessary to investigate to better understand the reason why no researcher used, in the analyzed texts, classification systems for the organization of audiovisual resources. Perhaps this is due to the fact that the classification systems were not designed specifically for digital resources, as the audiovisual resources in this scenario are understood, but it is a fact that classification systems are indispensable since they offer a theoretical and methodological ballast for solutions of problems concerning knowledge organization and the thematic information treatment.

The general proposal of this research was to identify which are and how the instruments of information organization are applied in the organization of audiovisual resources. It was noted, however, that there are still substantive issues, such as the audiovisual concept itself, which still require further investigation. Thus, we intend, in further studies, to undertake this task, with the use of conceptual and terminological analysis.

References

The references was made following the ABNT rules.

ALVARENGA, L.; SOUZA, R. M. F. A Universidade Federal de Minas Gerais no contexto do acesso aberto à informação científica: identificação de seus sistemas de informação. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 13., 2012, Rio de Janeiro. *Anais...*2012.Rio de Janeiro, 2012. p. 21.

ARAÚJO, W. T. Uso da informação audiovisual em bibliotecas: dados de pesquisas. *Informação & Sociedade*, João Pessoa, v.2, n.1, p.35-41, jan./dez. 1992.

ARQUIVO NACIONAL (Brasil). *Dicionário Brasileiro de Terminologia Arquivística*. Rio de Janeiro: Arquivo Nacional, 2005.

BARRETO, J. S. Desafios e avanços na recuperação automática da informação audiovisual. *Ciência da Informação*, Brasília, v. 36, n. 3, p. 17-28, set./dez. 2007.

BETHÔNICO, J. Signos audiovisuais e Ciência da Informação: uma avaliação. *Encontros Bibli*, Florianópolis, 2º número esp., 2º sem. 2006.

CEBRIAN-HERREROS, M. *Fundamentos de la teoría y técnica de la información audiovisual*. Madrid: Mezquita, 1983. 2 v.

CUNHA, M. B.; CAVALCANTI, C. R. O. *Dicionário de Biblioteconomia e Arquivologia*. Brasília: Briquet de Lemos, 2008. | 159

DIEUZEIDE, H. *Les techniques audiovisuelles dans l'enseignement*. Paris: Puf, 1965.

FAGUNDES, M. L. F.; PRADO, G. S. Videoteca digital: a experiência da videoteca multimeios do IA/UNICAMP. *Transinformação*, v. 11, n. 3, p. 293-299, set./dez. 1999.

FEATHER, J.; STURGES, P. (Eds.). *International Encyclopedia of Information and Library Science*. 2. ed. London: Routledge, 2003.

GONÇALVES, A. C. B. Os novos paradigmas da imagem em movimento. *DataGramaZero*, v. 3, n.1, fev. 2002.

MARTÍN-BARBERO, J.; REY, G. *Los ejercicios del ver: hegemonia audiovisual y ficción televisiva*. Barcelona: Gedisa, 1999.

MIRANDA, M. L. C.; GARBELINI, M. F. Tratamento técnico da documentação audiovisual na TV da Universidade Federal de Goiás. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 12., 2011, Brasília. *Anais...2011*. Brasília, 2011. p. 439-451.

MOURA, M. A. et al. Linguagens de indexação em contextos cinematográficos: a experiência de elaboração do tesouro eletrônico do cinema brasileiro. *Perspec. Ci. Inf.*, v.10, n.1, p. 54-69, jan./jun. 2005.

MUCHERONI, M. L.; RIBEIRO, C.; PAIVA, D. C. Descrição dinâmica para documentos multimodais em ambiente digital. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 13., 2012, Rio de Janeiro. *Anais...2012*. Rio de Janeiro, 2012. p. 17.

RAVAZI, F. C. *A organização dos recursos audiovisuais por meio das linguagens documentárias e das ontologias: aspectos teóricos e metodológicos*. 2014. 83 f. Trabalho de Conclusão de Curso (Graduação em Biblioteconomia) – Universidade Estadual Paulista, Faculdade de Filosofia e Ciências, Marília/SP, 2014.

SILVA, D. L.; SOUZA, R. R. Vocabulários e metadados para descrição de recursos multimídia: uma perspectiva de organização e representação do conhecimento no espaço digital. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 13., 2012, Rio de Janeiro. *Anais...2012*. Rio de Janeiro, 2012. p. 7.

SOUSA, E. E.; DUQUE, C. G. Recuperação da informação disponibilizada em vídeos: uma abordagem utilizando ontologia, processamento de linguagem natural e descritores multimodais. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 13., 2012, Rio de Janeiro. *Anais...2012*. Rio de Janeiro, 2012. p. 20.

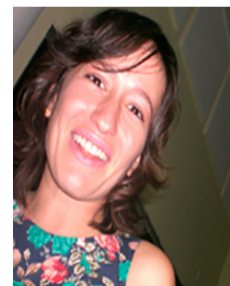
SOUSA, E. E.; DUQUE, C. G. Uso de ontologia para recuperação da informação disponibilizada em vídeos por meio de indexação multimodal. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 12., 2011, Brasília. *Anais...*2011.Brasília, 2011. p. 2730-2737.

SMIT, J. W. O documento audiovisual ou a proximidade entre as 3 marias. *Revista Brasileira de Biblioteconomia e Documentação*, São Paulo, v. 26, n. 1/2, p. 81-85, jan./jun. 1993.

Knowledge Organization and Information Design: a convergence study



Natalia Nakano
São Paulo State University
natinakano@gmail.com



Mariana Cantisani Padua
São Paulo State University
mariana.cantisani@gmail.com



Maria José Vicentini Jorente
São Paulo State University
mjorente@yahoo.com.br

Introduction

Systematizing knowledge, given the volume of information currently found, especially in digital environments, evidences the fields of Knowledge Organization (KO) and Information Design (ID), demonstrating interdisciplinary among them.

The subject, however, is still little discussed in Information Science (IS). KO, according to Hjørland (2008), can be understood in its narrow and broad senses. In the narrow sense, it encompasses traditional activities carried out in libraries, archives and memory institutions, such as classifying, indexing and cataloging and has, as its core disciplines, Librarianship, Documentation and Computer Science. In the broad sense, KO is related to the division of mental work, that is, KO as structure of areas for research, organizations of universities and other institutions of knowledge production and dissemination. Although Librarianship and IS are disciplines dedicated to KO in its narrow

sense, it is impossible to consider KO disregarding its broad sense, because, at present, other disciplines are also concerned with the study of knowledge and its organization. It can be affirmed, therefore, that KO exceeds the limits of the fields of knowledge and pervades disciplines never thought previously. Hjørland (2008, 2013) recognizes the intersection of KO with other fields such as Computer Science, Linguistics, Psychology,¹⁶² Social Sciences, Archival, Diplomatic and even Information Architecture, but does not mention the potential of Information Design (ID) and its contribution to KO.

Although ID has recently been consolidated as a discipline, its practical application traditionally concerns not only information organization and access, but it also includes an indispensable element in the socio-cognitive conception of KO: the human element. ID understands the individual inserted in a social and cultural context, and therefore extrapolates cognitive issues, especially after the informational explosion in digital information environments. According to Freitas, Coutinho and Waetcher (2013, p.2, our translation):

The ephemerality of information in the last decades, resulting from constant technological transitions, values, attitudes and needs occurred in society, turn the way of "treating information" an essential concern at the time of designing a project, and it doubles, when we discuss Information Design.

In this context, the problems and objectives of our research are outlined when presenting and discussing one of the points of convergence between the ID and KO: how have the two areas developed up to the present? How can ID be interrelated to KO systems in a way that makes them more efficient and attractive?

Among Hjørland's numerous valuable contributions to KO is the defense of the socio-cognitive view. The development of this view starts from the gaps perceived by Hjørland in user studies of the years 1970-1980. At that moment the classic model of Information Retrieval prevailed in IS (ROBERTSON, 1977). It is a linear and horizontal model, where information is constituted, on the one hand, as an instrument of transmission of signals (information as a thing) and, on the other hand, understood as a message that changes the structures of the receiver (cognitive concept of information).

Hjørland criticizes this approach, stating that KO is social, not just cognitive. The author defends, on the other hand, the Domain Analytical Approach (DA) as a more adequate perspective for domain analysis in KO. Formulated in the early 1990s as an alternative to the dominant cognitive view in Librarianship and

Domain analysis is a sociological-epistemological standpoint. The indexing of a given document should reflect the needs of a given group of users or a given ideal purpose. In other words, any description or representation of a given document is more or less suited to the fulfillment of certain tasks. A description is never objective or neutral, and the goal is not to standardize descriptions or make one description once and for all for different target groups. (Hjørland, 2008, p. 95)

According to the author, DA is the only approach to KO that examined epistemological issues in the field, comparing the assumptions made in different approaches to KO, and examining the issues of subjectivity and objectivity in KO. Important are collective visions, shared by many.

For Hjørland (2008), different points of view need different organizational systems. In the same way, any KO system is always biased toward a philosophical position that reflects the one who created it or commissioned it. There is no neutral platform from which knowledge can be organized. Information professionals need to develop a conceptual and practical approach that is consistent with the organization's objectives and values for which the system is developed. The socio-cognitive approach, beyond the cognitive questions, includes the social context in which the agent is inserted.

The author's main criticism related to the field of KO is that since 1950, we have seen KO migrate from one information system to another without developing a theoretical and epistemological basis for the development of these systems and platforms.

It can be argued that ID, as well as domain analysis approaches to the KO, evolved in parallel and in a similar way - by including the social view, hermeneutic circle and iterativity - that information systems should consider. ID can be understood as the science that seeks to prepare information for efficient and effective use in physical and digital environments, with the objectives of:

1. developing documents that are comprehensible, rapidly and accurately retrievable, and easy to translate into effective action.
2. designing interactions with equipment that are easy, natural, and as pleasant as possible. This involves solving many of the problems in the design of the human computer interface.
3. enabling people to find their way around in three-dimensional space with comfort and ease especially urban space, but also, given recent developments, virtual space). (HORN, 1999, p. 15).

We can assume that different approaches and paradigms coexist in all domains, and that ID emerges as a new science and a discipline that brings methodologies for the development of projects aiming to organize knowledge, structure information, and guide the realization of the search and information retrieval in an information system for the purpose of knowledge construction.

Information Interaction Models through Information Design

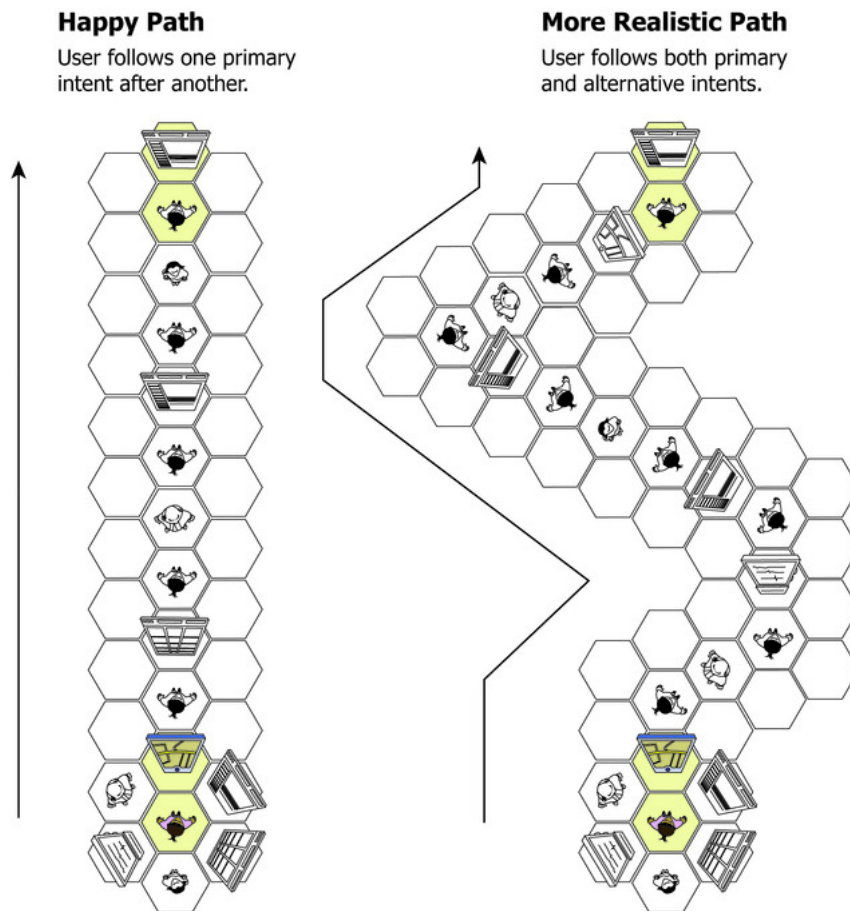
When considering KO, the question of the hermeneutical circle and its iterative functions of understanding and interpretation stands out here as a starting point. The basis of the hermeneutic circle is movement, in which the comprehension of the whole for the parts and the parts for the whole takes place. For this reason, in the present study, a correlation with the information search and knowledge-building models with ID is performed.

According to Palmer (1986), in phenomenological hermeneutics it is impossible to separate the object, the context and the agent; Gadamer (1997, p. 516, our translation) points out that it is necessary "to let things manifest as they are, without forcing them into our own categories. This implies a reversal of the usual direction. It is not we who point to what things are; on the contrary, things show themselves to us."

Planning/designing experiences of information interaction and use in informational environments that are effective requires, not only an awareness of the individual's cognitive characteristics but a clear understanding of how these agents will navigate the paths in the realization of this interaction with information, artifact and environment . As shown in Figure 1 (WILLIS, 2011), the path traveled by the agent that interacts with the information system does not happen directly from the subject to its objective. In this way, the designer has to consider that the agent, with its social and cultural differences inserted

in a given context, will take different paths. Thus, the efficient system must present different shortcuts for access to information, and must have resilient characteristics, adapting to the different needs of the agent.

Figure 1 - Projected Path vs. Traveled Path



Source: (WILLIS, 2011)

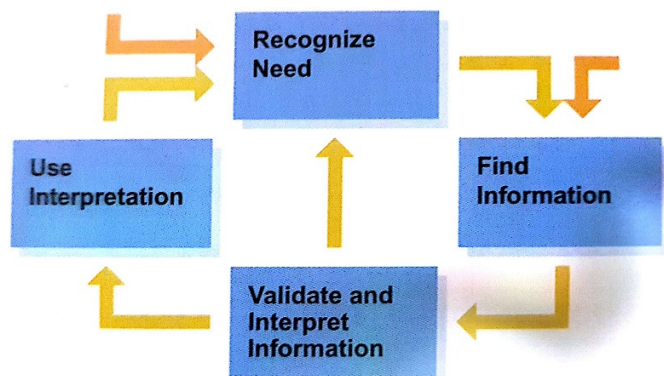
Already in the 1970s, Selle (1973) argued that the design of artificial environments should enable and facilitate the understanding of the meaning of objects and Information System by people. The designer, as a configurator agent, should then consider the dynamic reality of the subjects' interaction with the system, developing this process with the focus of creating a satisfying, pleasant and easy-to-understand experience.

The conception of interaction processes with information environments in ID has evolved from simple and static models to complex and dynamic models (RUSSEL-ROSE; TATE, 2013). Three models that shaped understanding in the process of access to information used by ID are presented in this study.

Norman (1988), from computing, was the pioneered research known as cognitive engineering. The model proposed by the author is concerned with the ways in which people interact with the machine; with the processes involved in the moment of interaction of the actors with the systems; with the interaction needs of individuals with digital artifacts to perform their tasks.

In the cognitive model, based on task performance, the agent creates a mental model of goals, performs a sequence of actions, executes actions, has a perception of the task, interprets and evaluates the results (Figure 2).

Figure 2 - Cognitive model proposed by Norman (1988)

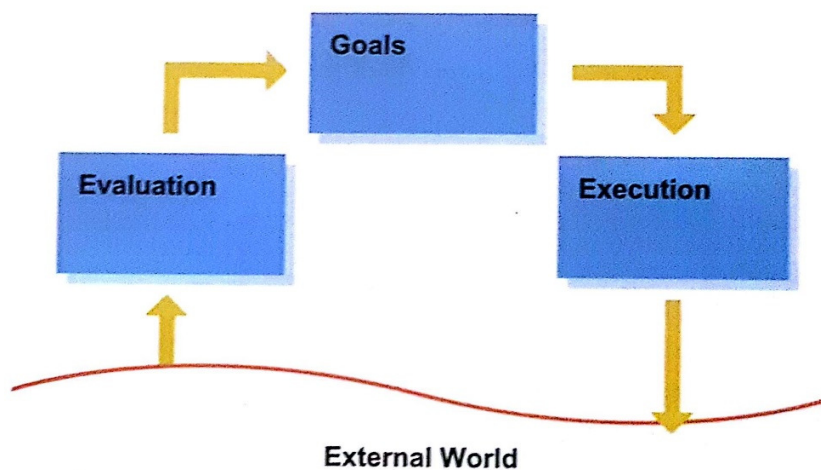


Source: (RUSSEL-ROSE; TATE, 2013, p.25)

It is fundamental, in this model by Norman (1988), that he recognizes the importance of knowledge domain: the greater the agent's knowledge about the domain, the greater the effectiveness of the queries, the accuracy and the relevance of the results (RUSSEL-ROSE; TATE, 2013.).

The model developed by Marchionini (1995), shown in Figure 3, known as the standard model, contrasted the classic IS model proposed by Robertson (1977). Marchionini (1995) argues that the search for information is a form of problem solving, involving a cycle of four activities: identification, articulation on the need for information, query formulation, and evaluation of results (STUCLIFFE and ENNIS, 1998). | 167

Figure 3 - Standard model proposed by Marchionini (1995)



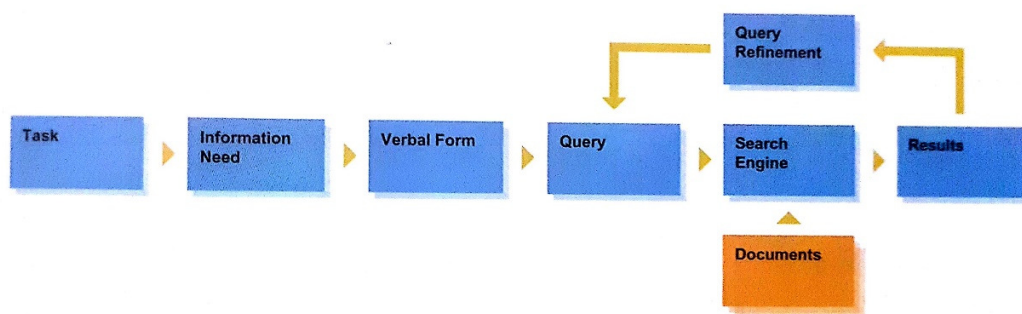
Source: (RUSSEL-ROSE; TATE, 2013, p.25)

Before engaging in carrying out the task, the person articulates his informational need through a verbal representation, which, in turn, results in a query. This query is then inserted into a document search tool from a collection. The tool presents a result, which, if not ideal, causes the individual to return to a new question.

The standard model is limited considering the dynamic relationships related to the human factors of interaction with information systems. In addition it does not consider that the need for information can be transformed according to the retrieved results. Moreover, the model does not consider individuals in their complexity: individuals are moved daily by feelings, impressions and feelings, which influence the forms of interaction with information, the informational environment and Information System.

The model proposed by Blandford and Attfield (2010), known as the Information Journey Model (Figure 4), was derived from empirical studies of informational behavior, and proposed the following framework: identification of information need; information search; interpretation and validation of information; and interpretation of usage.

Figure 4 - Information journey model proposed by Blandford and Attfield (2010)



Source: (RUSSEL-ROSE; TATE, 2013, p.25)

In this model, the information needs of the agents evolve according to the interaction with information and thus new goals are formulated as one gains knowledge in the domain. The search for information is, therefore, an iterative and interactive process in which the answer can change the question. It is worth noting that this model takes into account the serendipity phenomenon, represented by the gray arrows of Figure 4.

It can be observed that this model emphasizes the validation, interpretation and use of information - key activities, modeling of the evolution of information need, due to the multiple entry points of the model.

Currently, Information Design, in addition to considering the shared socio-cognitive vision with KO, has evolved to include User Experience (UX), the understanding of emotions that influence processes such as attention, recognition and memory generated by the interaction of the agents with information systems (DAMASIO, 2011; NORMAN, 2008).

UX investigates complexity issues such as subjectivity, uncertainty, transcendence and multidimensionality that extrapolate systems and their relationship with the agent,

investigates the context, culture and subjective issues related to users' emotions. Similarly to Hjørland (2008), considering not only the individual, but a group of individuals.

Conclusion

We concluded that, for an optimized systematization of knowledge, especially in digital environments, KO and ID are interdisciplinary areas of well-known convergence in conceptual, disciplinary and practical plans.

Both ID and KO consider that an efficient system structures on complex issues that include not only context and subject culture, but also uncertainties. In the context of uncertainty, an efficient system must be resilient for the satisfactory performance of tasks. In addition, the information system must operate intuitively, that is, it must present relevant affordance. Concepts, methods and theories widely discussed in ID which can optimize KO scope.

Thus, it is necessary to consider the complexity of the interrelationships of various disciplines with KO and the theoretical and practical contributions they can offer, such as interdisciplinarity with ID. For the academic scientific support of such an affirmation, this article presents conceptual models of ID that demonstrate a consonance with the knowledge produced in KO.

References

The references was made following the ABNT rules.

BLANDFORD, A.; ATTFIELD, S. *Interacting with Information*. Morgan & Claypool, 2010.

DAMÁSIO, A. *E o cérebro criou o homem*. São Paulo: Companhia das Letras, 2011.

FREITAS, R.F.; COUTINHO, S. G.; WAETCHER, H. N. Análise de Metodologias em Design: a informação tratada por diferentes olhares. *Estudos em Design | Revista (online)*, Rio de Janeiro, v. 21, n. 1, 2013. p.1-15.

HJORLAND, B. User-based and Cognitive Approaches to Knowledge Organization. *Knowledge Organization*, Wurzburg, v. 40, 2013. p. 11-27.

HJORLAND, B. What is Knowledge Organization (KO)? *Knowledge Organization*, Wurzburg, v. 35. n. 2/n. 3, 2008.

HORN, R. Information Design: Emergence of a New Profession. In: JACOBSON, R. (Org.). *Information Design*. Cambridge, Massachusetts: The MIT Press, 1999. p. 15-33.

GADAMER, Hans-Georg. A linguagem como *medium* da experiência hermenêutica. In: *Verdade e método I*. Trad. Flávio Paulo Meurer. Petrópolis: Vozes, 1997. p. 497- 631.

MARCHIONINI, G. *Information Seeking in Electronic Environments*. Cambridge: Cambridge University Press, 1995.

NORMAN, D. A. *The Psychology of Everyday Things*, New York: Basic Books, 1988.

| 170

NORMAN, D. A. *Design emocional: porque adoramos ou detestamos os objetos do dia-a-dia*. Rio de Janeiro: Rocco, 2008.

PALMER, R. *Hermenêutica*. Trad. Maria Luísa Ribeiro Ferreira. Lisboa: Edições 70, 1986.

ROBERTSON, S. E. Theories and models in information retrieval. *Journal of Documentations*, v. 33, n. 2, p 126-148, 1977.

RUSSEL-ROSE, T.; TATE, T. *Designing the search experience: the information architecture of discovery*. Morgan Kaufmann, 2013. 302 p.

SELLE, G. *Ideologia y utopia del diseño. Contribución a la teoría del diseño industrial*. Barcelona: Editorial Gustavo Gill, 1973.

SUTCLIFFE, A. G.; ENNIS, M. Towards a cognitive theory of information retrieval. *Interacting with Computers*, 10, 1998. p 321-351.

WILLIS, D. *Intent Paths*. 2011. Available at: <
<https://www.flickr.com/photos/uxcrank/sets/72157626618631035/with/5730905682/>> Access at:
5 mai. 2015.

Domain Analysis of the first ISKO-Brazil chapter editions



Exequiel Fontans
Universidad de la República
exequiel151974@gmail.com



Mario Barité
Universidad de la República
mabarite@gmail.com

L71

Introduction

Since its founding in 2007, ISKO-Brazil chapter has become the vehicle for the promotion and scientific exchange of Knowledge Organization in Brazil. In the first two editions of these events, according to the records, more than 80 papers were presented, which included 123 signatures from 38 institutions from 9 countries. In this context, it is important to analyze the domain of the production recorded in the proceedings of ISKO-Brazil chapter (Brasilia, 2011 and Rio de Janeiro, 2013), as this will allow the identification of the conditions under which scientific knowledge is constructed and socialized (Guimarães, Santos, Candido, & Pinho, 2014).

Smiraglia (2014, p.85) points out that "there has been some confusion about the definition of a domain, perhaps because the notion must embrace diversity. Some have written of discourse communities, disciplines, invisible colleges and even work ecologies, and all of these are kinds of domains." He concludes: "a domain is simply a group of scholars working on research problems that in some way perceived to be similar." (Smiraglia, 2014, pág. 85). This implies that they share an ontology, a set of common hypotheses, a working methodology that apply to the same research problems, and thus a common language that allows them to communicate (Smiraglia & Lee, 2012, p. 114)

Hjørland and Albrechtsen's (1995) Domain Analysis, postulated for Information Science, is a central methodological approach in Knowledge Organization, subdivided into eleven methodological approaches (Hjørland, 2002). Among them, we find bibliometric

studies, especially mentioning citation and cocitation analysis (Oliveira & Grácio, 2013). For this work, in addition to citation analysis, the analysis of co-occurrence of words, and of collaboration are used. These bibliometric approaches are enhanced by combining them with the analysis of social networks. Among the numerous antecedents that apply the bibliometric approach and social network analysis, we can mention the characterization of the research in Information Science in Brazil reported by Liberatore and Herrero-Solana (2013), the analysis of GT 7 from the 13th ENANCIB in 2012 (Guimarães, Moreira González, & Alencar, 2012) and more recently the work by Guimarães et al. (2014), in which they analyze the proceedings of the congresses in Brazil, Spain and North America for the 2011-2013 period.

In this context, the present proposal aims to contribute to domain analysis of the thematic field Knowledge Organization in Brazil through the combined use of bibliometric tools and social network analysis to the documentary corpus formed by the proceedings of the congresses organized by the Brazilian ISKO chapter. In particular, it is intended to: a) identify the addressed themes, b) represent personal, institutional and cross-country collaboration, and c) analyze and visualize the citations used.

Methodology

To carry out this study, the following methodological procedures were followed:

1 - Obtaining the documentary corpus

The documentary corpus consisted of the two conference proceedings from ISKO-Brazil chapter that took place in Brasilia in 2011 and Rio de Janeiro in 2013 (Guimarães & Dobedei, 2012, 2013).

Data from the papers were fed into a spreadsheet. The collected data were: paper title, keywords, authors, affiliation, year, bibliographic references.

As the papers had no keywords, these were obtained through an indirect method: the significant words in the papers' titles representing their content were identified and selected.

2 - Treating the corpus.

In this stage, the authors' names were normalized, the titles in other languages (9 in English and 17 in Spanish) were translated into Portuguese, and the names of the cited authors were normalized.

3 - Data Analysis

The matrices of keywords, authors, and cited authors, in addition to the matrices of attributes necessary for further processing were made with UCINET 6 and NetDraw 2 software (Borgatti, Everett, & Freeman, 2002).

Analyses of social network analysis metrics were performed through centrality measures. This concept was proposed by Freeman (1979), with the contributions of Newman (2000, 2001) and are very relevant for application to collaboration networks. The main measures of centrality are briefly described below.

Degree Centrality: This measure is the most common and intuitive of the centrality measures: it seems reasonable to infer that nodes with more relationships with others (the better connected) tend to be more influential in the network.

Betweenness Centrality: if a node is in the shortest path (geodesic distance, in theory of graphs) between another pair of nodes is sure to have influence between the information that circulates between that pair of nodes.

Closeness Centrality: this measure in a network provides an indicator of how close a node is, on average, to the rest of the network nodes. Nodes with a high index of proximity access the information circulating on the network very fast. If the network is collaborative, authors with a high index of closeness can be seen as authors who access very fast the information circulating on the network and in turn the information they generate arrives very fast to other authors of the network (Newman, 2000, 2004, 2010; Ye, Li, & Law, 2011). If it is a network of keywords, closeness can be interpreted as thematic affinity.

For citation analysis, we start from the presumption that two or more documents that are cited together contain a thematic similarity, at least for the citing one. In a corpus constituted by the proceedings of a disciplinary congress such as the one addressed in this paper, it is expected that this theme will coincide with the topics of the congress. Therefore, it is assumed that the highly cited documents represent a thematic nucleus that

could be used to identify and visualize the relationships between citation authors (Miguel, Moya-Anegón, & Herrero-Solana, 2006).

4 - Processing Products

- A. General data.
- B. Keyword analysis.
- C. Collaboration analysis.
- D. Citation analysis

Results

a - General data

Table 1 presents the basic data of the collaboration network. Some relevant data are highlighted below. Out of the 83 documents that compose the corpus, 43% (n = 36) are signed by one single author. Out of the documents made in collaboration, 64% (n = 30) are signed by two authors. On the other hand, only 18% (n = 22) of authors do not have collaborative works. These data explain that the average number of authors per document is slightly higher than two (2,15).

Table 1. Basic data from collaboration network

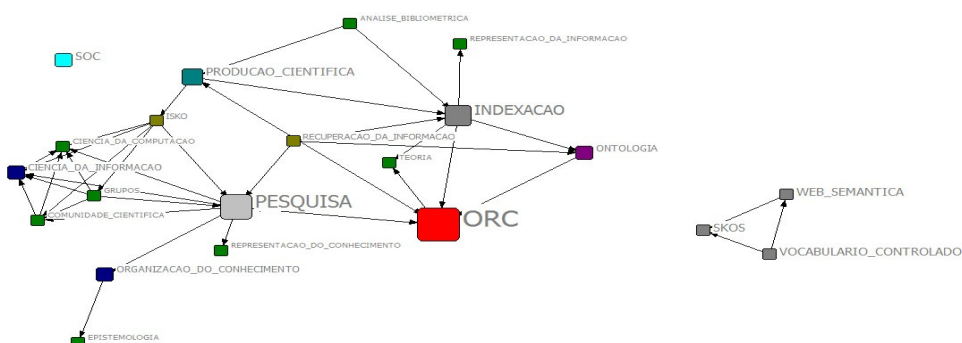
Total number of documents	83	
Documents by one single author	36	43%
Documents in coauthorship	47	57%
Documents by two authors	30	64%
Documents by three authors	10	21%
Documents by more than three authors	7	15%
Total number of authors	123	
Authors with no collaboration	22	18%
Authors with some collaboration	101	82%
Authors with one single paper in collaboration	84	68%
Authors with two papers in collaboration	17	14%
Authors with three or more papers in collaboration	0	0%
Total number of collaborations	101	
Mean documents per author	0.67	
Mean authors per documents in collaboration	2.15	
Total number of authors with three or more papers in collaboration	0	0%

Source: research data.

b - Keyword Analysis

The analysis of the papers titles resulted in 130 keywords. To facilitate visualization, the selected keywords were the ones higher than 5 (Figure 2). Table 2 shows the centrality measures of the 21 words that meet this requirement. The words are ordered by intermediation (Betweenness); it can be said that the seven keywords that appear first in this ranking are the hard core theme of the congress of the Brazilian chapter of ISKO (PESQUISA; ORC; INDEXACAO; ORGANIZACAO DO CONHECIMENTO; PRODUCAO CIENTIFICA; ISKO; RECUPERACAO DA INFORMACAO). Meanwhile, proximity (Closeness) indicates the words that are more connected with the rest of the thematic (SKOS; WEB SEMANTICA; VOCABULARIO CONTROLADO; SOC; EPISTEMOLOGIA; REPRESENTACAO DA INFORMACAO; ANALISE BIBLIOMETRICA; REPRESENTACAO DO CONHECIMENTO; TEORIA; ONTOLOGIA).

Figure 2. Keywords Network of degree greater than 5.



Source: research data.

Table 2. Centrality Measures of keywords network of degree greater than 5.

Nombre	Rank	Grado	Grado Normalizado	Rank	Cercania	Cercania Normalizada	Rank	Intermediación	Intermediación Normalizada
PESQUISA	1	9	0,155	13	703	0,195	1	65	0,117
ORC	4	5	0,217	12	706	0,199	2	25	0,154
INDEXACAO	2	7	0,116	9	711	0,192	2	25	0,083
ORGANIZACAO_DO_CONHECIMENTO	7	2	0,062	7	716	0,181	3	15	0,052
PRODUCAO_CIENTIFICA	5	4	0,085	10	710	0,188	4	14	0,023
ISKO	3	6	0,047	11	708	0,184	4	14	0,008
RECUPERACAO_DA_INFORMACAO	6	3	0,047	11	708	0,188	5	12	0,021
ONTOLOGIA	6	3	0,070	7	716	0,187	6	0	0,036
CIENCIA_DA_INFORMACAO	4	5	0,062	8	712	0,182	6	0	0,021
CIENCIA_DA_COMPUTACAO	4	5	0,039	8	712	0,179	6	0	0,000
COMUNIDADE_CIENTIFICA	4	5	0,039	8	712	0,179	6	0	0,000
GRUPOS	4	5	0,039	8	712	0,179	6	0	0,000
SKOS	7	2	0,047	1	764	0,117	6	0	0,000
WEB_SEMANTICA	7	2	0,047	1	764	0,117	6	0	0,000
VOCABULARIO_CONTROLADO	7	2	0,039	1	764	0,117	6	0	0,001
ANALISE_BIBLIOMETRICA	7	2	0,039	4	720	0,177	6	0	0,017
TEORIA	7	2	0,039	6	717	0,186	6	0	0,018
EPISTEMOLOGIA	8	1	0,039	2	731	0,171	6	0	0,006
REPRESENTACAO_DA_INFORMACAO	8	1	0,039	3	726	0,176	6	0	0,025
REPRESENTACAO_DO_CONHECIMENTO	8	1	0,039	5	718	0,179	6	0	0,012
SOC	9	0	0,062	1	774	0,118	6	0	0,003

Source: research data.

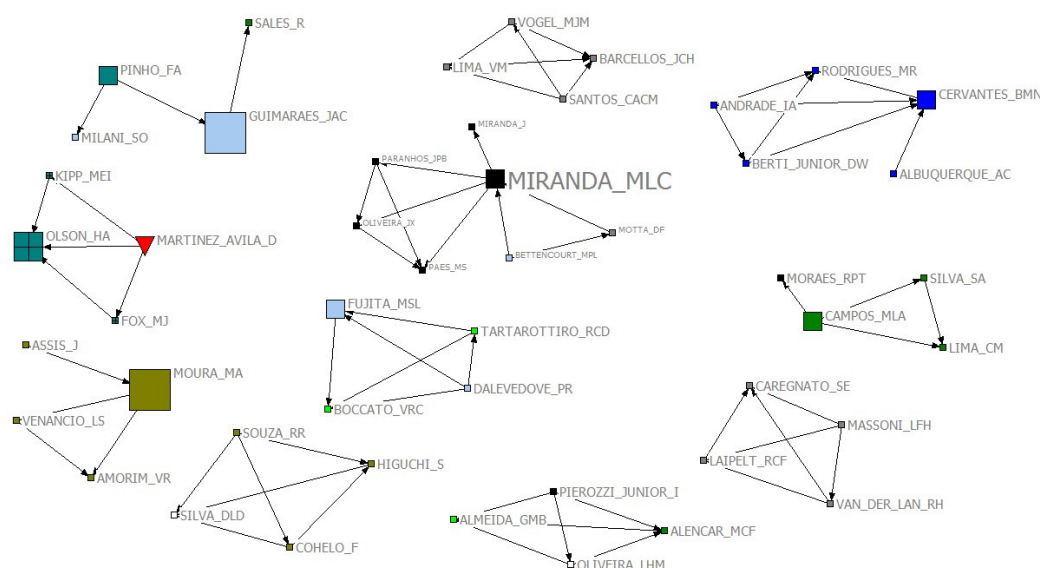
c - Collaborative analysis

Figure 3 shows the collaboration groups formed by at least 4 members. We observe 9 groups of 4 members, one of 5 (Cervantes BMN) and one of 7 (Miranda MLC). We show here three of the four most productive authors (Guimarães, Moura and Olson)

that are listed in Table 3. In addition, the collaboration between different institutions (color of the nodes) and countries (form of the nodes) is observed. Out of the 11 verified groups, 7 have some kind of inter-institutional collaboration. Figure 4 shows the authors without collaboration and groups with 2 and 3 members.

Table 4 shows the number of authors per country; as it is natural, 83% of the participants are from Brazil, followed by the US and Spain with 5% and 4%, respectively. The authors are distributed per institutions, as shown in Table 5; 50% of the authors come from 7 Brazilian institutions: UNIRIO; USP; UNESP; UFMG; UFF; UEL and UFRGS.

Figure 3. Collaboration network formed by groups of 4, 5 and 7 members.



Source: research data. Note: The size of the node and its tag represent the number of oral communications presented; the shape represents the countries; the color represents the affiliation.

Table 3. Detail of the authors with more contributions.

Nombre	Cant. Art.	Colab.	País	Filiación
GUIMARAES_JAC	4	2	BRASIL	UNESP
MOURA_MA	4	2	BRASIL	UFMG
OLSON_HA	3	2	EEUU	UWM
BARITE_M	3	2	URUGUAY	UDELAR

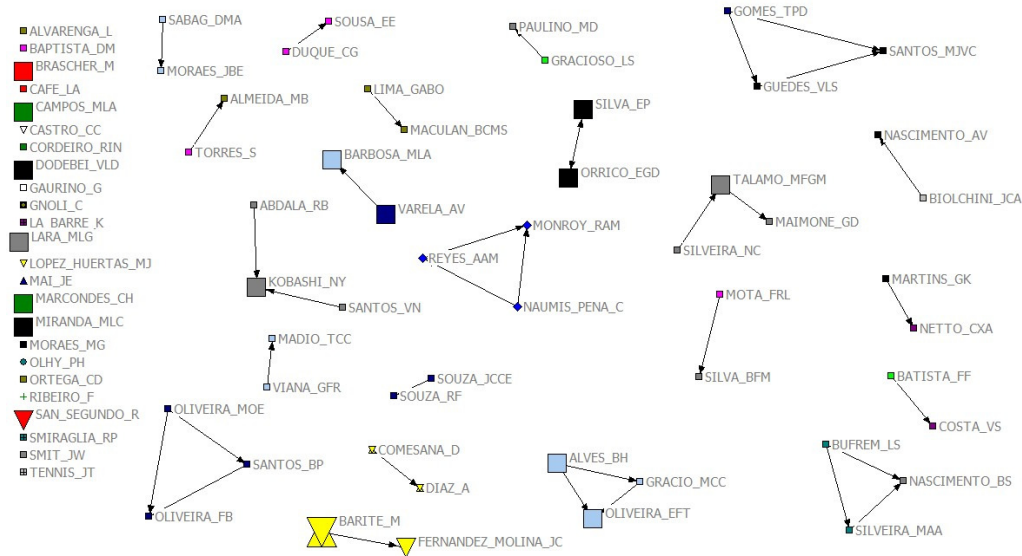
Source: research data. Note: Cant. Art.= Number of articles; Colab.= Collaborations.

Table 4. Details of participating countries.

País	Cantidad	Porcentaje
ALEMANIA	1	1%
BRASIL	102	83%
DINAMARCA	1	1%
EEUU	6	5%
ESPAÑA	5	4%
ITALIA	1	1%
MEXICO	3	2%
PORTUGAL	1	1%
URUGUAY	3	2%
Total	123	

Source: research data.

Figure 4. Collaboration network formed by authors without collaboration and groups with 2 and 3 members.



Source: research data. Note: The size of the node and its tag represent the number of oral communications presented; the shape represents the countries; the color represent the affiliation.

Table 5. Detail of participating institutions.

Institución	Cant. Inst.	Cant. Aut.	Sub total	Porc.	Porc. Acum.
UNIRIO	1	13	13	11%	11%
USP	1	12	12	10%	20%
UNESP	1	11	11	9%	29%
UFMG	1	9	9	7%	37%
UFF	1	6	6	5%	41%
UEL; UFRGS	2	5	10	8%	50%
UFPA; UFSCAR; UNIVERSITY OF WISCONSIN	3	4	12	10%	59%
FGV; IBICT; UDELAR; UFPE; UNAM; UN	6	3	18	15%	74%
EMBRAPA INFORMATICA					
AGROPECUARIA; UC3M; UFC; UFES; UFRJ; UFSC; UGR	7	2	14	11%	85%
ASSEMBLEIA LEGISLATIVA DO ESTADO DE MINAS GERAIS; CNEN; EMBRAPA MEIO AMBIENTE; IPEA; ISKO; THE UNIVERSITY OF WASHINGTON INFORMATION SCHOOL; UCB; UFAL; UFBA; UFPB; UNIVERSIDAD DE SALAMANCA; UNIVERSIDADE DO PORTO; UNIVERSITA DE PAVIA; UNIVERSITY OF COPENHAGEN; UNIVERSITY OF ILLINOIS AT URBANA CHAMPAIGN	15	1	15	12%	98%
Sin Dato	1	3	3	2%	100%

Source: research data. Note: Cant. Inst.= Number of institutions; Aut.= Number of authors; Porc = Percentage; Porc. Accum = Cumulative percentage.

d - Citation analysis

For citation analysis, the number of citations for the first author of the cited work was counted (without taking into account the other authors in collective papers); the same criterion was established for the authors of presented papers: the first author of the paper was considered "citation author" for those contributions of collective authorship.

Table 6 presents the details of the most cited authors. Five authors, 1% (Hjorland B, Dahlberg I, López-Huertas MJ, Fujita MSL and Smit JW), concentrate 9% of the citations (n = 74). If we consider the authors who received at least 2 citations (22%, n = 112), they represent half of the reported citations.

As analyzing this volume of information (512 authors, 799 citations) is not very productive as a criterion of cut, we took those authors who received at least 3 citations (n = 53), representing the 10% most cited of the corpus, with 35% of total citations. From the

authors of papers with their citations, we took only those presenters who had cited this nucleus of 53 most cited authors. A matrix with the 85 authors obtained from this operation was then performed. Of these authors, 23 are lone citations (no citations received), 37 are lone cited and the rest are cited and citation. It is highlighted that 60 papers cited some of these authors of this hard core. This implies that 83% of the papers (on the 72 papers that have bibliography), cited at least one of the 53 authors of the most cited nucleus. ¹⁸¹

Table 6. Citation Details.

Total general	Cant. Aut.	Acum	Porc. Acum.	Cant. Citas	Total citas	Citas Acum.	Porc. Acum.
HJORLAND B	1	1	0%	24	24	24	3%
DAHLBERG I	1	2	0%	23	23	47	6%
LÓPEZ-HUERTAS MJ	1	3	1%	10	10	57	7%
FUJITA MSL	1	4	1%	9	9	66	8%
SMIT JW	1	5	1%	8	8	74	9%
BARITÉ M; BEGTOL C; CAMPOS MLA; GARCÍA GUTIÉRREZ AL; GARDIN J-C; GUIMARAES JAC	6	11	2%	7	42	116	15%
BOOK INDUSTRY STUDY GROUP; CAPURRO R; DODEBEI VLD; ISO; OLSON HA; SMIRAGLIA RP	6	17	3%	6	36	152	19%
FOSKETT AC; GUIZZARDI G; LANCASTER FW; MILLS J; RANGANATHAN SR	5	22	4%	5	25	177	22%
FOUCAULT M; GIL LEIVA I; GUARINO N; KOBASHI NY; LA BARRE K; MAI JE; MIRANDA MLC; OHLY HP; ORRICO EGD; SVENONIUS E; TENNIS JT	11	33	6%	4	44	221	28%
AITCHISON J; BUCKLAND MK; CASTELLS M; CHAUMIER J; CLARKE SGD; FARRADANE JEL; GARCIA MARCO FJ; GNOLI C; GONZALEZ DE GOMEZ MN; INGWERSEN P; ISKO; MARTÍNEZ-ÁVILA D; MOREIRO GONZÁLEZ JA; NOVAK JD; OLMEDA GÓMEZ C; PINHO FA; SHATFORD LS; SMITH B; SOERGEL D; TÁLAMO MFGM	20	53	10%	3	60	281	35%
Varios autores	59	112	22%	2	118	399	50%
Varios autores	400	512	100%	1	400	799	100%

Source: research data. Note: Cant. Aut.= Number of authors; Acum= Accumulated; Porc. Cumulative= Accumulated Percentage; Cant Cit.= Number of citations per author; Citas Acum. = Accumulated citations.

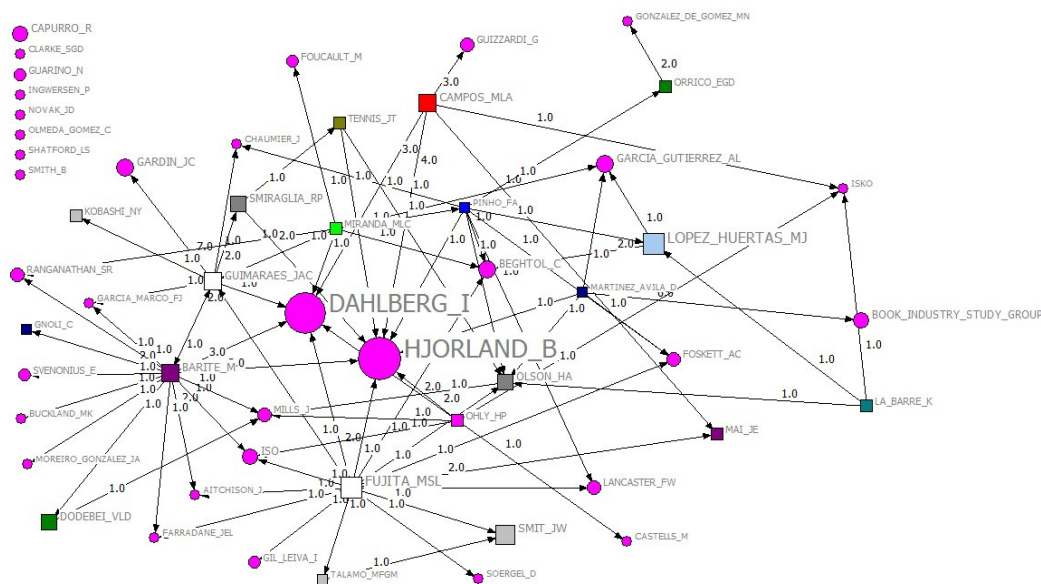
In Figure 5, we present the citation network of the 53 most cited authors of the corpus. A first aspect to consider is that the giant component of the network includes all the authors who participated in ISKO-Brazil chapter congresses. This indicates that for each node of the giant component there is a path that relates to any other node in the network. In other words, this figure shows the configuration of a community of local² researchers who share a citation network that includes the international referents of the discipline.

² In a broad sense, given that 7% of the authors are from Brazil.

In Figure 6 the authors who did not participate in the analyzed conference were eliminated. The result is a citation network that includes in its giant component 19 of the 20 authors belonging to the nucleus of most cited authors. This reinforces the idea expressed before: the existence of a local community that is cited by the rest of the participants of the conferences, but also citing one another. | 182

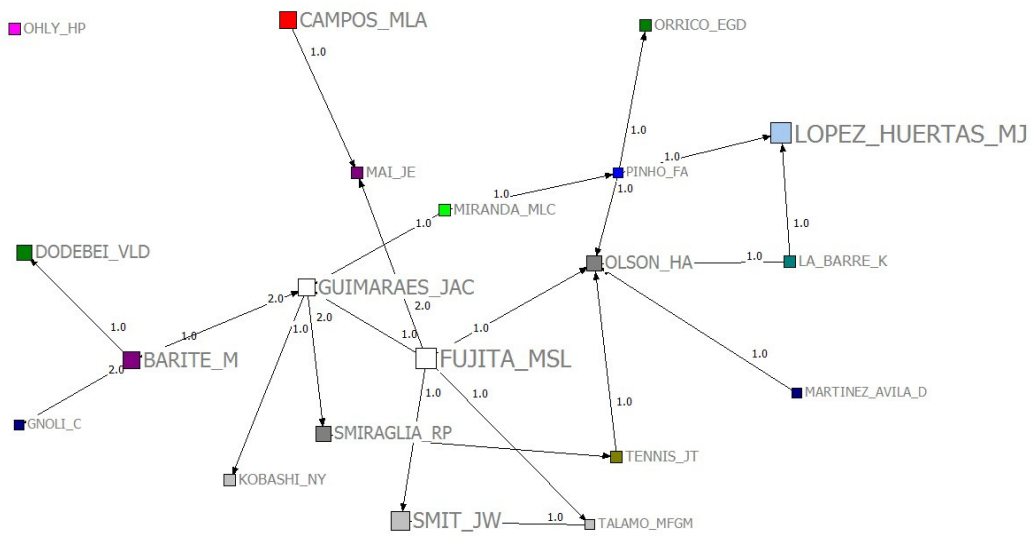
Figure 7 includes the network of authors from Figure 6 who cite the core of the most cited authors. The colors of Figures 6 and 7 reveal high interagency collaboration.

Figure 5. Citation network of the 53 most cited authors of the corpus.



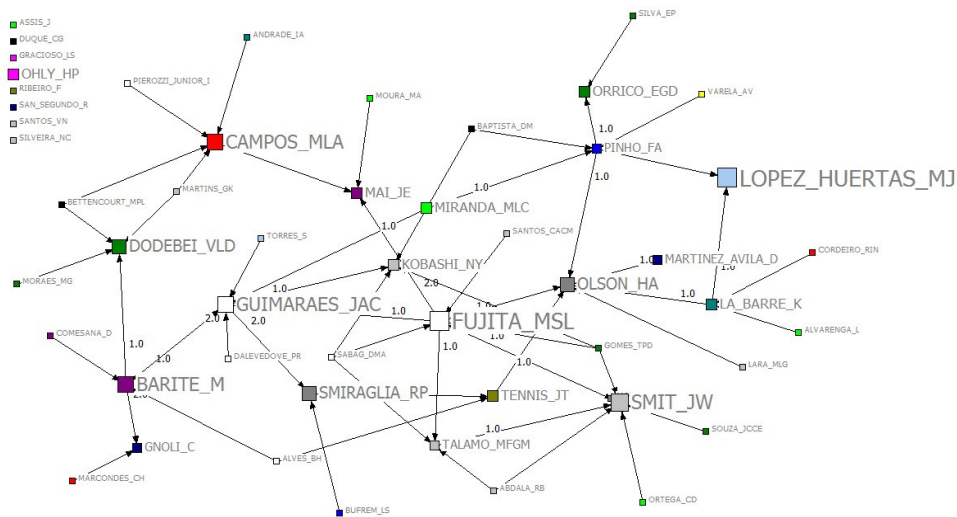
Source: research data. Note: The size of the node and its tag represent the number of citations received, including self-citations; the tip of the arrow points to the cited author and the number, the number of times he quoted it; the circles represented the authors who did not present papers and the squares represent those who did; the colors of the squares identify institutions.

Figure 6. Citation network of the authors who participated in ISKO-Brazil chapter congresses belonging to the nucleus of most cited authors.



Source: research data. See note in figure 5.

Figure 7. Citation network of the authors who participated in ISKO-Brazil chapter congresses.



Source: research data. See note in figure 5.

Discussion and conclusions

This paper analyzed the proceedings of the congresses of ISKO Brazilian chapter

in its two editions, 2011 and 2013, through the use of bibliometric techniques and social networks analysis. Bibliometric studies constitute a coherent and objective approach to analyze and characterize a scientific domain (Hjørland, 2002; Oliveira & Grácio, 2013); however, its results should be taken with caution if they are not combined with qualitative methods. In this context, this work is exploratory and seeks to describe a complex phenomenon with a valid tool, but with intrinsic limitations. ¹⁸⁴

We identified a core made up of seven keywords that represent the themes addressed in these two events, namely: PESQUISA; ORC; INDEXACAO; ORGANIZACAO DO CONHECIMENTO; PRODUCAO CIENTIFICA; ISKO; RECUPERACAO DA INFORMACAO; and another nucleus of keywords with a high index of closeness, indicating that they are better connected with the rest of the words of the study (SKOS; WEB SEMANTICA; VOCABULARIO CONTROLADO; SOC; EPISTEMOLOGIA; REPRESENTACAO DA INFORMACAO; ANALISE BIBLIOMETRICA; REPRESENTACAO DO CONHECIMENTO; TEORIA; ONTOLOGIA). The fact that the keywords were obtained in an indirect way (from the titles of the presentations) is a limitation that must be taken into account.

The method of co-publication analysis identified the collaborative networks that allowed to visualize inter-institutional and international collaboration. The sample of the documentary corpus did not allow to identify communities with this method. Co-authorship, understood as the joint signature of a paper, is still one of the most used methods to study collaboration, however, these approaches have their limitations that have been reported, among others, by Katz and Martin (1997).

The citation analysis³ allowed to identify and visualize a community of authors that make up the nucleus of most cited authors in the analyzed documentary corpus. This nucleus is made up of referents of the discipline at international level and local authors, which could be accounting for the important development of KO in Brazil. In addition, at local level, there is an inter-institutional collaboration, although this part of the analysis was not deepened to be able to describe it more precisely.

³ Considering the size of the documentary corpus for this work, self-citations were not excluded, which definitively influence the result of the analysis.

References

Borgatti, S. P., Everett, M. G., & Freeman, L. C. (2002). *Ucinet 6 for Windows: Software for Social Network Analysis* (p. 302). Harvard: Analytic Technologies.

Freeman, L. C. (1979). Centrality in Social Networks Conceptual Clarification. *Social Networks*, 1(3), 215-239.

| 185

Guimarães, J. A. C., & Dobedei, V. [Organizadores]. (2012). *Desafios e perspectivas científicas para a organização e representação do conhecimento na atualidade* (p. 285). Marília: ISKO-Brasil; FUNDEPE.

Guimarães, J. A. C., & Dobedei, V. [Organizadores]. (2013). *Complexidade e organização do conhecimento: desafios de nosso século* (p. 310). Rio de Janeiro: ISKO-Brasil; FUNDEPE.

Guimarães, J. A. C., Moreiro González, J. A., & Alencar, M. F. (2012). A análise documental no universo científico dos ENANCIBS: elementos para uma análise de domínio. En *XIII Encontro Nacional de Pesquisa em Ciência da Informação - XIII ENANCIB 2012* (p. 17). Rio de Janeiro.

Guimarães, J. A. C., Santos, A. Y. dos, Candido, G. G., & Pinho, F. A. (2014). A dimensão temática da pesquisa em organização do conhecimento: uma análise de domínio dos congressos nacionais e regionais da ISKO (Brasil, Espanha e América do Norte). *Scire*, 20(2), 19-25.

Hjørland, B. (2002). Domain analysis in information science: Eleven approaches – traditional as well as innovative. *Journal of Documentation*, 58(4), 422-462. doi:10.1108/00220410210431136

Hjørland, B., & Albrechtsen, H. (1995). Toward a new horizon in information science: domain-analysis. *Journal of the American Society for Information Science*, 46(6), 400-425. doi:10.1002/(SICI)1097-4571(199507)46:6<400::AID-ASI2>3.0.CO;2-Y

Katz, J. S., & Martin, B. R. (1997). What is research collaboration? *Research Policy*, 26(1), 1-18. doi:10.1016/S0048-7333(96)00917-1

Liberatore, G., & Herrero-Solana, V. (2013). Caracterización temática de la investigación en Ciencia de la Información en Brasil en el período 2000-2009. *Transinformação*, 25(3), 225-235. doi:10.1590/S0103-37862013000300005

Miguel, S., Moya-Anegón, F., & Herrero-Solana, V. (2006). El análisis de co-citas como método de investigación en Bibliotecología y Ciencia de la Información. *Investigación Bibliotecológica*, 21(43), 139-155.

Newman, M. E. J. (2000). Who is the best connected scientist? A study of scientific coauthorship networks. En *Complex Networks* (Vol. 370, pp. 337-370). doi:10.1103/PhysRevE.64.016132

Newman, M. E. J. (2001). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences of the United States of America*, 98(2), 404-9. doi:10.1073/pnas.021544898 | 186

Newman, M. E. J. (2004). Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences of the United States of America*, 101(1), 5200-5205. doi:10.1073/pnas.0307545100

Newman, M. E. J. (2010). *Networks: An introduction* (p. 772). New York: Oxford University Press.

Oliveira, E. F. T., & Grácio, M. C. C. (2013). Studies of Author Cocitation Analysis: a approach for Domain Analysis. *IRIS: Informação, Memória e Tecnologia*, 2(1), 12-23.

Smiraglia, R. P. (2014). *The Elements of Knowledge Organization* (p. 101). New York: Springer.

Smiraglia, R. P., & Lee, H. L. (Eds.). (2012). *Cultural Frames of Knowledge* (p. 158). Würzburg: Ergon-Verlag.

Ye, Q., Li, T., & Law, R. (2011). A coauthorship Network Analysis of Tourism and Hospitality Research collaboration. *Journal of Hospitality & Tourism Research*, 37(1), 51-76. doi:10.1177/1096348011425500

Teaching in knowledge organization and representation: languages and standards



Cibele Araújo Marques dos Santos
University of São Paulo
cibeleac@usp.br



Eduardo de Abreu de Jesus
University of São Paulo
eduardo.jesus@usp.br



João Ricardo de Luca
University of São Paulo
joao.luca@usp.br

Introduction

The tools used to organize and represent the information contained in printed or digital documents make up the knowledge organization systems, consisting of documentary languages such as controlled vocabularies, lists of subject headings, classification systems and thesauri, as well as other types of schemas: taxonomies and ontologies (CARLAN and BRÄSCHER, 2011). Such systems, especially documentary languages, according to Campos (2006), can be defined by the control of vocabulary and by the established dynamics of semantically related terms, generically covering a specific domain of knowledge.

These languages encompass theoretical and methodological contributions that include linguistics, logic, indexing, terminology and the construction of controlled

thesauri and vocabularies, as well as the use of classification systems and lists of subject headings. The Internet and Semantic Web allowed the emergence of formats and standards such as XML schemas, SKOS (Simple Knowledge Organization System) and ontologies.

Thus, the interdisciplinarity and complexity of these tools makes it imperative to study these resources and rethink the teaching of documentary languages and other systems of knowledge organization by information professionals.

In a project developed to identify teaching practices and study the trends of the formats, standards and norms of these systems, it was possible to observe that the literature on knowledge organization and representation presents a gap in the research on teaching-learning methodologies of languages and schemes of information organization.

Methodology

The bibliographic research was carried out in the international databases of Information Science LISA and ISTA. We also searched the multidisciplinary databases Web of Science and SCOPUS, but no articles were found. In national databases, BRAPCI was searched, where only 3 articles on teaching indexing were found and in SCIELO, with no results. The search strategy used keywords such as "documentary languages", "bibliographical classifications", "thesaurus" and "controlled vocabulary", crossing with "teaching" in Portuguese and English, in the title, text and abstract and in LISA using the thesaurus with similar strategy. Some papers were selected, mainly from LISA base for analysis.

Characteristics of Knowledge Organization Systems

The teaching of knowledge organization tools such as documentary languages demands theoretical and practical interdisciplinary contributions that present a certain complexity, considering the characteristics of existing systems and schemes.

For Leise (2008), who presents a well-didactic article, a controlled vocabulary differs from an indexing list by several factors: time, form and purpose of elaboration, and the essential characteristics of a controlled vocabulary are the relationships between its terms, such as the control of synonymy, the hierarchical structure and the associative relation. The author adopts the terminology used by ANSI/NISO Z39.19 (Controlled Vocabulary) to refer to documentary languages and how the standard includes taxonomies in this scope (NATIONAL INFORMATION STANDARDS ORGANIZATION, 2010). It also alerts to the terminological diversity regarding languages and the difficulty of defining and conceptualizing the terms of the area. The article defines and explains using tables, in addition to the controlled vocabulary, the synonyms ring, authority control, taxonomy and thesauri. It also brings the concept of facet and its importance to the categorization and structuring of a controlled vocabulary.

Hedden (2008) also didactically defines the concepts of Controlled Vocabulary, as a list of terms for the elaboration of indexing and categorization, being controlled because it defines the term to be used in indexing and its importance is more easily visible when regarding Information retrieval. For the author, thesaurus includes, in addition to the concept of controlled vocabulary, a hierarchical structure that provides information about each term and its relationships. The author considers taxonomy as a controlled vocabulary with hierarchical structure, but exempt from equivalence and synonymy relations and other requirements of the traditional thesaurus.

Navigation taxonomy for Leise (2008) corresponds to a controlled vocabulary to organize websites or other digital content, creating tags for navigation categories and different levels of hierarchy that must be designed to reflect a particular context, as it is already the case of vocabularies, and should be used in conjunction with website content management systems.

While the design and construction of controlled vocabularies have been presented at a technical level in detail over the last decades, the methodological level has been neglected, so researchers have argued that the elaboration of these

languages needs to be based on the analyzes and understandings of contexts in which they are inserted (MAI, 2008).

Information and guidance on the management of knowledge organization systems are important, since the literature on the maintenance and revision of documentary languages and controlled vocabularies is scarce and dispersed, and there are also few studies on the teaching of this activity.

The maintenance of a controlled vocabulary, as Leise (2008) emphasizes, should be constant, considering addition/deletion/revision of terms, reorganization of hierarchies, modification, addition and deletion of the facet tag, updating, synonyms control, changes in the control process and definition of vocabulary property.

In addition, Leise (2008) points out that although there are many new concepts and ideas to be learned in the field, many indexers consider the elaboration of a controlled vocabulary as an extension of their indexing activity. The construction of controlled vocabularies is related to indexing, according to Hedden (2008), in relation to its objectives. For the author, the skills needed for indexing are the same skills needed for developing languages and vocabularies, and these skills include deciding on the best word for a concept, creating general and specific term in a hierarchy, defining the main entry, and the preferred/non-preferred terms, reference and relationship between terms.

The question here is how to develop these skills so that undergraduates and graduate professionals can learn the skills needed to create and manage knowledge organization systems. There is also a need to learn how to select documentary languages when necessary, make languages compatible, develop and/or use software and formats, learn international standards for controlled vocabularies and schemes for interoperability.

Teaching and standards

The researched literature has provided articles on the typology of documentary languages, their definitions and structure, uses, applications and norms, but few papers on the teaching of these languages.

For Caro-Castro (2006), the important in documentary languages is what is lexically coded, not grammatically, which is why we eliminate the flexed forms of noun terms chosen for descriptors, so these are units of languages with a meta-representation function and are no longer ambiguous because they have been standardized. This means that the relations between concept and descriptor were controlled. For this, methods that establish relations of equivalence are used, and the meaning of the descriptors is specified through codes and reach notes. The author also recalls that the organization of a knowledge field in a thesaurus, for example, favors its understanding by indicating not only represented concepts, but the interrelationships that unite them.

Esteban-Navarro (1995) reinforces, in a paper presented at a conference on the teaching of documentary languages, that librarians should have knowledge of principles for maintaining these languages in the information units and develop these competences in progress with the presentation of the methodology for applying the basic terminology principles for thesaurus management, including scientific fundamentals and practical exercises. He considers as necessary theoretical contributions: terminology, lexicographical and terminological studies in specific area, sources of terminographic work of localization, identification and follow-up of new terms and concepts through search, selection and retrieval of technical and scientific documentation, current and explicit, attribution of classification codes to the concepts of terminological records, management of conceptual systems by the use and management of documentary languages, standardized terminologies through the implantation of scientific information systems, publication and diffusion of secondary documents and control of the diffusion of terms through bibliometric studies.

The presence of documentary languages in undergraduate studies, according to Espelt (1995), who presented a paper in the same conference, is not limited to the definition, basic characteristics and practical application of the most used

languages, but should also focus on terminology and facet analysis. For the author, it is important that professionals are familiar with theories, principles and terminological guidelines. The student must learn to be aware that his/her function is based on the existence of several levels of language and his/her job is to move across the terminology of documents, the terminology of information system and the terminology of users.

The normalization of terms facilitates treating documents, Esteban-Navarro (1995) recalls, and favors the progressive homogeneity of vocabulary, retrieval and diffusion of information, allowing authors, information professionals and users to use the same terms to designate the same concepts. Moreover, the analysis of how the definitions of concepts and the control of terminological and conceptual systems can be applied to the management of documentary languages, since it provides vocabularies based on the principles of structuring and conceptual relation, that communicate the precise situation of a concept within a thematic field. Documentary language appears as normalizing at the morphological, syntactic and semantic levels to guarantee stability and efficiency in information organization.

The author exposes the importance of teaching the definition of terminological concepts and their elaboration rules based on ISO 704-1987. The learning of the techniques for definition construction allows to acquire a fundamental resource to maintain and to construct these languages, since the establishment of relations between the terms must be preceded by restricted fixation of the extension of the concepts. In order to obtain univocity, several techniques of disambiguation or elimination of the synonymy or polysemy are used, which, according to Espelt (1995), include the logical operations among concepts and the operations applied in the definition are instruments to determine synonymy and polysemy among terms, to establish specificity of the language and correct application.

Also in controlling the relationship among concepts, the method of constructing concept systems is the best strategy in Esteban-Navarro's (1995) understanding in order to know the exact meaning of the concept, since they present themselves in relation to others within the conceptual field, and as the basic foundation of logical-semantic structures of documentary languages that reflect the

systematization of existing knowledge in a given field, according to its developmental state and according to the dominant taxonomy trend. The author presents the analysis of the main criteria for structural grouping or construction of taxonomies to show how the generic and ontological relations are the hierarchies, also indicating the different systems of graphical representation that can be used.

It is important to teach the application of these types of relationships in concrete cases to show the three types of systems of documentary concepts: hierarchical, partitive and mixed, and grouping criteria of concepts that use themes, disciplines or facets, as well as the rules for presenting terms in thesauri according to international standards (ESTEBAN-NAVARRO, 1995).

For Espelt (1995) one must also consider that knowledge about the characteristics of scientific and technical texts is part of the preparation for the analysis of the documentary content. In addition, she considers the knowledge of concepts systems and the methods applied in the definition of terms for the correct application of documentary languages in indexing and retrieval of information to be appropriate. The author reminds that the team elaborating a thesaurus should be composed of specialists in the thematic area, terminologists and documentalists. Likewise, she emphasizes the importance of developing exercises with students such as analysis of scientific-technical texts, definitions of terms aiming to become descriptors, lists of terms with relationships and elaboration of concept systems. In addition, faceted classifications are important for the analysis of documentary content and also for indexing quality, and she reinforces that the literature in this area of education is scarce and indexing manuals are very general about this, so that the teaching of documentary and indexing languages should have specific guidelines.

Moreiro-Gonzalez (1995) points out that, at the same time, there is a significant number of students grammatically underprepared or even with difficulties to interpret the text they analyze. Therefore, it is proposed that learning take place in phases, initially with the introduction of preparatory language studies, the teaching of classification systems, indexing techniques and abstracting, having, in this phase, a predominant application component. In a second phase, deepening of the theme

must occur, with the inclusion of a variety of elements and knowledge that allow representative transformations in the process.

Clarke and Zeng (2012) showed that the principles for constructing a thesaurus have changed dramatically from ISO 2788 to ISO 25964, published in 2011, which we consider to be justified by the worldwide changes with information technology, with the need for interoperability for Semantic Web, where the distinction between term and concept is fundamental, an important contribution implemented in the new standard. Thus, SKOS (Simple Knowledge Organization System), W3Consortium's recommendation was developed for the publication of vocabularies and thesauri on the Web.

In comparison to other solutions, SKOS offers an alternative whose application is simple and fast (PASTOR-SÁNCHEZ; MARTÍNEZ-MÉNDEZ; RODRÍGUEZ-MUÑOZ, 2012). The authors inform that SKOS is formally defined as an ontology that allows to represent any type of knowledge organization system through RDF. Its scope extends to practically all controlled vocabularies: classifications, thesauri, subject headings, taxonomies, glossaries, etc. In SKOS, the elements of a vocabulary are represented by concepts among which semantic, hierarchical, associative, language, synonymy and alternative access to a concept are established through labels.

Conclusions

This research allowed to identify a gap that in the literature, especially in Brazil, about methodology and teaching practices of documentary languages, the importance of the theme for teachers and professionals, as well as the specificity of the knowledge necessary for the performance in the area, both for use and for the elaboration of documentary languages.

The literature review resulted in papers on the definition, methods of elaboration and management of documentary languages, and theoretical and practical contributions for teaching, formatting and use of knowledge organization systems. Issues related to the skills that need to be taught, the interdisciplinary and

trends for systems development and interoperability that need to be included in the courses was shown.

Due to the relevance of the issues related to teaching in the area, we consider the importance for further research, analyzing the literature and experiences to propose ways to improve teaching-learning processes and skills development.

References

The references was made following the ABNT rules.

CAMPOS, M. L. A., GOMES, H. E. Metodologia de elaboração de tesouro conceitual: a categorização como princípio norteador. *Perspectivas em Ciência da Informação*, v.11, n.3, 348-359, 2006.

CARLAN, E.; BRÄSCHER, M. Sistemas de Organização do Conhecimento na visão da Ciência da Informação. *RICI: Revista Ibero-americana de Ciência da Informação*, v. 4, n. 2, p. 53-73, ago./dez. 2011.

CARO-CASTRO, Carmen. Léxico y documentación: del lenguaje natural al lenguaje documental. In: ROMAN, Maria Teresa Fuentes; REY, Jesus Torres del. *Nuestras palabras: entre el léxico y la traducción*. Frankfurt: Veuver, 2006. p. 127-145.

CLARKE, Stella G. Dextre; ZENG, Marcia Lei. From ISO 2788 to ISO 25964: the evolution of thesaurus standarts towards interoperability and data modeling. *ISQ* n. 1, v. 24, p. 20-26, 2012.

ESPELT, C. Bases teóricas em la enseñanza de lenguajes documentales. I Conferencia ISKO- Espanha, Madrid, 4-5 Novembre, 1993. *Proceedings*. Saragosa, 1995.

ESTEBAN-NAVARRO, M. A. Aplicaciones de la Terminologia para la docência de la gestión de lenguajes documentales. I Conferencia ISKO- Espanha, Madrid, 4-5 Novembre, 1993. *Proceedings*. Saragosa, 1995.

HEDDEN, H. Controlled vocabularies, thesauri and taxonomies. *The Indexer*, v. 26, n. 1, p. 33-34, 2008.

LEISE, F. Controlled vocabularies: an introduction. *The Indexer*, v. 26, n. 3, p. 121-126, 2008.

MAI, Jens-Erik. Actors, domains, and constraints in the design and construction of controlled vocabularies. *Knowledge organization*. v. 35, n. 1, p. 16-29, 2008.

MOREIRO-GONZÁLEZ, José Antonio. La enseñanza de los lenguajes documentales en las diplomaturas de Biblioteconomía y Documentación: problemática y tendencias. I Conferencia ISKO- Espanha, Madrid, 4-5 Novembre, 1993. *Proceedings*. Saragosa, 1995.

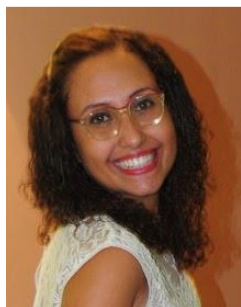
MORGAN, J.; BAWDEN, D. Teaching knowledge organization: educator, employer and professional association perspectives. *Journal of Information Science*, v. 32 p. 108, 2006.

NATIONAL INFORMATION STANDARDS ORGANIZATION. *Guidelines for the construction, format, and management of monolingual controlled vocabularies*, 2010. (ANSI/NISO Z39.19-2005 R2010.)

PASTOR-SÁNCHEZ, J.; MARTÍNEZ-MÉNDEZ, F.; RODRÍGUEZ-MUÑOZ, J. Aplicación de SKOS para la interoperabilidad de vocabularios controlados en el entorno de linked open data. . *El profesional de la información*, v. 21, n. 3, p. 245-253, mayo-junio 2012.

Acknowledgement: Universidade de São Paulo. Programa Ensinar com Pesquisa.

Genealogy of the concept of Information Science in Brazil: a discursive analysis from founding journals in the area



Larissa de Mello Lima
São Paulo State University
larissalima.unesp@gmail.com



João Batista Ernesto de Moraes
São Paulo State University
jota@marilia.unesp.br



Daniel Martínez-Ávila
São Paulo State University
dmartinezavila@marilia.unesp.br

Introduction

The concept of Information Science emerged in Brazil in the 1970s in a university context and as a product of discursive practices and transformations. To present the genealogy of the concept of Information Science in Brazil, it is fundamental to carry out a prior characterization of the agents responsible for shaping the emergence of the area in a national context.

The journal *Ciência da Informação* is considered one of the most important journals of the area and was created in 1972 by the former IBBD, currently IBICT. In the same year, the institute encouraged UFMG School of Librarianship to create *Revista da Escola*

de Biblioteconomia da UFMG. This initiative by IBICT is a reflection of the creation of the first master's degree course in Information Science in 1970. (MENEZES; SILVA; COZINET, 2007).

It is possible to affirm that the historical conjuncture of Information Science in Brazil has a structured foundation under institutional bias. Instances of scientific validation of Information Science in Brazil are denominated the institutions that characterize and even determine the history of the area in the country such as IBDD and the journals cited above. ¹⁹⁸

Delineating a discourse of the area possesses characteristics evidently peculiar to the scientific discourse (or the discourse that seeks to be scientific), with traits that render the analysis little fluid, by the very performance of the ideology.

It can be understood that the scientific discourse, in which Information Science is allied, there is a constant redefinition and repetition of previous positions and displacements provoked by the action of power (research institutions, researchers, fostering agencies, universities).

It can be said that it is part of the scientific validation movement itself, that is, as an author X of smaller "impact" cites author Y of greater impact and recognized by the scientific community (power), author X seeks binding to the point of "overcoming" author Y, in this way it is possible to say that scientific practice is hostage to its enunciative formulations.

There is, in the analyzed articles, a "game" that refers to the ideological action and institutions that legitimize the performance of this ideology, insofar as the discourse is an instance that records this performance.

That is to say, within the scope of the discursive tissue, it is important to situate that it is on the performance of the subject that the discourse is constructed, but such action is conditioned to the direct action of the ideology that is translated in the multiplicity of positions that subjects can assume in the discourse thus falling within the institutional framework.

Methodology

The methodology is based on the theoretical methodological assumptions of French matrix of discourse analysis (FOUCAULT 1986; 2010, PÊCHEUX, 1975)

Discourse analysis, in the context of Information Science, has already been presented or used in several theoretical and practical studies (for example, FROHMANN 1992; 1994a; FROHMANN 1994b; 2001; BUDD & RABER 1996; BUDD 2006; RADFORD 2003; RADFORD & RADFORD 2005; HAIDER & BAWDEN 2007; CAMPBELL 2007; 2011; OLSSON 2010; MARTÍNEZ-ÁVILA 2012).

In this study, we analyze articles selected from the journals *Ciência da Informação* and *Revista da Escola de Biblioteconomia da UFMG* from the 1970s, the birth period of Information Science in Brazil, as both are configured as instances of scientific validation of the area in the national context.

Regarding the selection of articles, it is possible to draw some conclusions that provide an outline of some characteristics inherent to scientific discourse. The first conclusion that deserves attention is the fact that the object of analysis is inserted in the typology of the scientific text; this, in turn, is composed by scientific communities, which are also discursive communities as they legitimize certain enunciative instances to the detriment of others.

Regarding the method, the construction of the observation device brings some elements that, when articulated in the articles during the analysis, will delineate the conceptual and discursive course of the discipline. Placing this analysis as archaeological means describing how discursive practices are situated at a given point.

Archaeological discourse analysis is not intrinsically concerned with what ought to be, in the sense of reaching the ultimate resolution to a puzzle; it is concerned with discursive practices as they are at a point in time. (BUDD, 2006, p. 73).

To describe them, the structure of Pêcheux (1975) reviewed by Orlandi (2000) is taken up again and it is outlined in stages guided by fundamental questions.

- "First treatment of surface analysis "

Moment when one has a primary contact with the linguistic surface of the text. It is also at this moment that the element of the archive is exposed, that is, the corpus that will be submitted for analysis.

- "Transformation of the linguistic surface into discourse object"

To effect this transformation, it is necessary to ask a guiding question: "What is said in this discourse? What is said in another discourse? From this structuring, the discursive object is exposed from the discursive linguistic phenomena (paraphrase, polysemy, polyphony) that affect it.

- "From the discursive object to the discursive process":

Moment when the guiding question is: "Why this and not another one?". At the answer in each analysis the discursive process will be reached, which shows the relation that said has with its exterior.

Results

Table 1 shows the concepts of IS extracted from the articles selected from the journals *Ciência da Informação* and *Revista da Escola de Biblioteconomia da UFMG* in the 1970s.

Table 1: Comparative table across concepts

ARTICLES	CONCEPT OF IS
"From Bibliography to Information Science: History and Position"	Scientific discipline that carries out investigations.
"Course on Information Science for students of Librarianship"	Science that refines Librarianship process, systems and phenomena.
"Bibliometric Relations Between the Research Front and Literature Reviews: a Study applied to Information Science"	.Discipline that concerns communication phenomena.

“Las ciencias de la información en la escuela de bibliotecología”	Interdisciplinary Science that investigates information properties and behavior.
“Distribution of Brazilian Geological Literature: a Bibliometric Study”	Science based on Librarianship underpinnings.
“Post Graduation in Librarianship.”	Science that aids Librarianship teaching.
“Futurology - Childhood Illness of Librarianship”	Discipline purely scientific.
“School of Information Science: Analytical bibliography.”	Interdisciplinary science.
“Bibliometric behavior of Portuguese Language as a vehicle for representing information”	Interdisciplinary science as it relates, in this case with Linguistics and Bibliometry.
“Planning a teaching unit in Information Science.”	Interdisciplinary science.
“Education in Information Science in the 1980s.”	Interdisciplinary science that relates professional practice with scientific research.
“Post-Graduation in Librarianship and Information Science: Reflection, Suggestions and Experiences”	Interdisciplinary science.

Source: by the authors

It is noteworthy that, over the years the articles were published, IBBD, one of the scientific validation instances of Information Science in the country, had as its main objective, the development in Science and Technology to fill a gap in the national system of scientific and technological development, as already emphasized by Menezes; Silva; Couzinet (2007).

It is believed that this institution has an approximation with the line of thought of Vannevar Bush that confers to Information Science the character of science that arises to

solve the question of exponential increase of the postwar scientific production. That is, IBICT begins to reflect, from 1976, what was advocated by Bush in 1945, which would be to pay attention to the technological perspective of data processing.

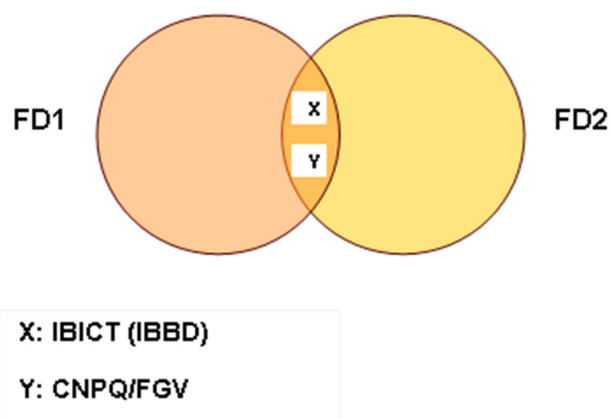
In contrast to the theoretical ideals of the institution, in search of an effective valuation of the technological perspective, it is possible to notice, through the comparative table, the revelation of two discursive formations focusing directly on the area. ²⁰²

The notes by Smit, Tálamo and Kobashi (2004) converge for the characterization of discursive formation 1, thus illustrating that the question of interdisciplinary is properly discursive, as it follows the formation of the constitution of science based on post modernity without actually containing a clear orientation on what this interdisciplinary means.

As discursive formation 1 (FD1) there is the characterization of Information Science from its interdisciplinary. As discursive formation 2 (FD2) there is the understanding that Information Science is a recent discipline that lacks theoretical underpinnings.

At the point of intersection between the two discursive formations (X, Y) we have IBICT and the CNPq/FGV that are configured as instances of scientific validation of Information Science, as IBICT was the body responsible for the creation of both journals, which had their articles analyzed. And this body has a direct relationship with CNPq and FGV because of the pioneering initiatives that culminated in its creation. That is, these are factors that, within a political conjuncture, form the institutionalization perspective of the area.

Figure 2 - Discursive formations of Information Science



Source: by the authors

Conclusion

We concluded that the genealogy of the concept of Information Science in the area counts on the presence of two discursive formations in the 1970s. In FD1, there is the perspective focusing on its interdisciplinary. In FD2, the fact that the area is recent requires more scientific research that consolidates its theoretical goals.

It is understood, from the perspective of this study, that the historical-discursive institutionalization of the area in Brazil in the 1970s forms the genealogy of the concept of Information Science in the country, being structured through discursive formations that now seek to sediment the question of interdisciplinary that focuses on the area, now emphasize the fact that Information Science is recent, requiring well-defined theoretical underpinnings.

References

The references was made following the ABNT rules.

BUDD, J. Discourse Analysis and the Study of Communication in LIS. *Library Trends*, v. 55, n. 1, p. 65-82, 2006.

BUDD, J.; RABER, D. Discourse Analysis: Method and Application in the Study of Information. *Information Processing & Management*, v. 32, n. 2, p. 217-226, 1996.

CAMPBELL, D. G. The Birth of the New Web: A Foucauldian Reading of the Semantic Web. *Cataloging & Classification Quarterly*, v. 43, n. 3-4, p. 9-20, 2007.

CAMPBELL, D. G. RDA and RDF: A Discourse Analysis of Two Standards of Resource Description. In: SMIRAGLIA, Richard P. (Ed.) *Proceedings from North American Symposium on Knowledge Organization*, Vol. 3, 2013, p. 207-216. | 204

FOUCAULT, M. *A arqueologia do saber*. Rio de Janeiro: Forense, 1986.

FOUCAULT, M. *A ordem do discurso*. São Paulo: Edições Loyola, 2010.

FREITAS, L. S. *Na Teia dos Sentidos: análise do discurso da Ciência da Informação sobre a atual condição da informação*. 2001. Tese (Doutorado em Ciência da Informação e Documentação) - Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo, 2001.

FROHMANN, B. The Power of Images: A Discourse Analysis of the Cognitive Viewpoint. *Journal of Documentation*, v. 48, n. 4, p. 365-386, 1992.

FROHMANN, B. Discourse Analysis as a Research Method in Library and Information Science. *Library and Information Science Research*, v. 16, p. 119-138, 1994a.

FROHMANN, B. Communication Technologies and the Politics of Postmodern Information Science. *Canadian Journal of Information and Library Science*, v. 19, n. 2, p. 1-22, 1994b.

FROHMANN, B. Discourse and Documentation: Some Implications for Pedagogy and Research. *Journal of Education for Library & Information Science* v. 42, n. 1, p. 13-26, 2001.

HAIDER, Jutta; BAWDEN, David. Conceptions of "information poverty" in LIS: a discourse analysis. *Journal of Documentation*, v. 63, n. 4, p. 534-557, 2007.

MARTÍNEZ-ÁVILA, D. Problems and Characteristics of Foucauldian Discourse Analysis as a Research Method. In: SMIRAGLIA, Richard P.; Lee, Hur-Li (Eds.), *Cultural Frames of Knowledge*. Würzburg: Ergon, 2012. p. 99-110.

MENEZES, E. M.; SILVA, E. L.; COUZINET, V. A Ciência da Informação na França e no Brasil. *Datagramazero*, v.8, n. 6, 2007.

OLSSON, M. R. Michel Foucault: Discourse, Power/Knowledge, and the Battle for Truth." In: LECKIE, Gloria J.; GIVEN, Lisa M.; BUSCHMAN, John E. (Eds.) *Critical theory for library and information science: exploring the social from across the disciplines*. Santa Barbara: Libraries Unlimited, 2010, p. 63-74

ORLANDI, E. L. P. *Discurso e leitura*. 5. ed. São Paulo: Editora da Unicamp, 2000.

PÊCHEUX, M. *Semântica e discurso: uma crítica à afirmação do óbvio*. Tradução de Eni P. Orlandi [et al.]. Campinas: Editora da UNICAMP, 1975.

RADFORD, G. P. Trapped in our own discursive formations: toward an archaeology of library and information science. *The Library Quarterly*, v. 73, n. 1, p. 1-18, 2003.

RADFORD, G. P.; RADFORD, Marie L. Structuralism, post-structuralism, and the library: de Saussure and Foucault. *Journal of Documentation*, v. 61, n. 1, p. 60-78, 2005.

SMIT, J. W.; TÁLAMO, M. F.; KOBASHI, N. A determinação do campo científico da ciência da | 205
informação: uma abordagem terminológica. *Datagramazero*, v. 5, n. 1, fev. 2004.

The Applied Dimension of Knowledge Organization

A Discourse Analysis of the Semantic Web



207

D. Grant Campbell
University of Western Ontario
gcampbel@uwo.ca

This presentation will examine the Semantic Web as a discourse that rests on a set of unspoken political and social assumptions that distinguish it from traditional bibliographic control on the one hand, and from Big Data and social media systems on the other. Using Foucault's concept of the discursive formation, together with Charles Taylor's concept of the "social imaginary," this presentation will argue that the Semantic Web operates on a principle of shared responsibility and assumptions of honesty that preserve a link between the purpose of data collection and its ultimate use.

The Semantic Web has been around nearly as long as the Web itself, and has undergone numerous transformations in its conceptions and its implementations, the most recent being the rise of linked data. Tim Berners-Lee originally envisioned it as a ubiquitous network of data rather than of documents: a network that would mine the resources of the Deep Web—data contained in Web-accessible relational databases, and other structured contexts—to create systems that were capable of performing sophisticated inferences, checking for data quality and validity, and assembling data to respond to specific questions and needs (Berners-Lee, Hendler, & Lassila 2001). In its first imagined incarnations, the Semantic Web looked much like the data-saturated world in which we now live. Human beings went about their daily lives doing much of what they'd always done, but accompanied by intelligent agents that would relate ongoing activities to a data infrastructure, prompt human beings with suggestions, and proactively adjust the environment according to pre-established requirements.

This infrastructure would rest upon a series of standards established and maintained by the World Wide Web Consortium, including the Resource Description Framework (RDF), the RDF Schema, the Working Ontology Language (OWL), and Friend of a Friend (FOAF). Since the Consortium has no legislative or coercive function, no one forces an institution or a developer to use these standards. But since these standards address such ubiquitous processes as information description, classification, vocabulary control and social networking, the Consortium encourages organizations to opt in to these standards, thereby making the world gradually more interoperable, releasing data from its many silos to be combined and recombined for various purposes.

This utopian vision of a data infrastructure has two important features. First it is created by the actual users and creators of the data, rather than by a designated intermediary such as a library: users and creators who consciously decide to use Semantic Web standards. Second, it grows from a core of deliberate and purposeful statements. RDF is based on the concept of the triple—subject, predicate and object—in which a specific community of data creators and users expresses consensus about the attributes of a resource and its relationships with other resources. We have, therefore, a community of data creators, system designers and information users taking it upon themselves to make statements about data. As Michel Foucault reminds us, these statements, like all discursive formations, have power relationships embedded into them. By their very existence, they form images of the conditions under which statements emerge; the authorities that determine which statements can or cannot be made and by whom; and the grids of classifications that group and differentiate various statements (Foucault 2010).

These discursive formations enable us to clarify how Semantic Web practices differ from those of library systems on the one hand, and those of Big Data on the other. Libraries have traditionally assigned trained intermediaries, in the form of cataloguers and indexers, to create their own discursive formations in the form of catalogue records, classification systems, indexes and thesauri. End users have traditionally had to abide by the categories and descriptions created by these intermediaries.

Big Data systems, on the other hand, have a very different view of the surfaces of emergence. In the case of Big Data, what matters is not the statement, but the trace.

Uses of information typically leave traces that, when collected, aggregated and studied for patterns, enable organizations to derive insights and predict outcomes that are often unrelated to the original data use. Big data for targeted advertising and recommender systems might well study the time of hovering over a Netflix option, and the kinds of programs enjoyed on specific devices; Big Data for predictive analysis might use search queries, geographic location and social media behaviour. In these cases, what matters is not the statement—the purposeful articulation of data characteristics for the purposes of meeting specific needs—but the lingering trace of these statements, and the accumulation of these traces into something that bears only an incidental relationship to the data’s original purpose.

Charles Taylor suggests that in our modern life and culture we are haunted by “social imaginaries”: inherited assumptions and theories that were formulated long ago, and which now lie half buried and only partially recognized (Taylor 2005). If we take his concept and apply it to the discourse of the Semantic Web, compared to that of traditional libraries and to Big Data, we find an intriguing parallel with concepts of political economy that emerged in the seventeenth and eighteenth centuries. Libraries work on an assumption that owes much to Thomas Hobbes’s *Leviathan*. Information requires a social contract, very similar to Hobbes’s social contract, in which information users sign over their autonomy to a designated body: in so doing, they are saved from a chaotic and brutal state of nature (Hobbes 1985). Big Data works on a principle similar to that of Adam Smith’s invisible hand: the notion that individual acts of self-interest, when aggregated together, achieve a social good that was unrelated to the selfish motivations of the individual acts (Smith 2000, 485).

The Semantic Web, however, echoes Jean-Jacques Rousseau’s concept of the Social Contract: an arrangement whereby a community voluntarily takes control of itself, and by sacrificing freedom in a superficial sense achieves a more profound freedom (Rousseau 1988, 24). The creators of the Semantic Web envision an environment in which enlightened data users, data creators, and data system designers voluntarily choose to adopt a set of mutually agreed-upon standards at some immediate cost in autonomy, speed and cost, in order to create a stronger mutual empowerment within and between knowledge domains and information communities.

References

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The Semantic Web. *Scientific American* May 2001: 35-43.

Foucault, M. (2010). *The archaeology of knowledge and The discourse on knowledge*. New York: Vintage. | 210

Hobbes, T. (1985). *Leviathan*. Middlesex: Penguin.

Rousseau, J-J. (1988). *On the social contract*. Indianapolis: Hackett.

Smith, A. (2000). *The wealth of nations*. New York: Modern Library.

Taylor, C. (2005). *Modern social imaginaries*. Durham: Duke University Press.

Semantic Web Technologies applied to Knowledge Organization: SKOS for the construction and use of decentralized controlled vocabularies



José Eduardo Santarem Segundo
University of São Paulo
santarem@usp.br



Caio Saraiva Coneglian
São Paulo State University
caio.coneglian@gmail.com

Introduction

The creation of Tim Berners-Lee's Semantic Web project has emerged in the face of the difficulties in locating, describing and retrieving information in Web environments, so it is directly related to all kinds of science, discipline or area that copes with these processes.

The idea of the Semantic Web arises from the need for computational agents to interact with systems without the need for human intervention. This requires the use of a large amount of technologies and concepts.

The Semantic Web is characterized as a project that aims to provide the sites with descriptive and thematic information for users, as well as information that can be processed and identified by computers, automatically. Thus, it would be a way of making information available to machines/software as well as information for users (BERNERS-LEE, LASSILA, HENDLER, 2001).

The Semantic Web gives structure to the meaning of the Web page, creating an environment conducive for search agents to perform sophisticated tasks and deliver them to the user (Bersers-Lee, Lassila, Henderson, 2001).

The challenge of the Semantic Web has been, each day, to provide a language capable of expressing data and rules at the same time, in order to allow the deduction of

new data and rules from any knowledge representation system to be imported or exported on the Web.

The technologies and concepts that involve Semantic Web have, as a fundamental point, the creation of a new structure of data storage. The main point is in separating the presentation and the structure from the available content, treating the atomic units of information as independent components. ²¹²

This separation will allow information retrieval in several ways, regardless of the search, once you know the structure of the data. This new information retrieval format should facilitate the association of information and help minimize the problem of using the same information in multiple systems.

Although the Semantic Web is presented as a seven-layer structure, over the years it has been showing us that the technologies and languages that make up these layers can be applied individually or in clusters, in order to build digital environments that allow information storage and retrieval with meaning, that is, using semantic relations.

In the context of information organization, it is understood that these technologies can be used in the construction and use of controlled vocabularies, in order to facilitate the decentralization and automation of their use.

One of the problems in this research is the difficulty that organizations have in transforming documentary languages into independent digital objects that can be used in a unified and integrated way by various types of digital platforms. The main examples of platforms that require controlled vocabularies are library catalogs, digital repositories, and electronic scientific journals. There are also a lot of other platforms and information services that could use controlled vocabularies.

The need for constant use of documentary languages in these environments is latent and the processes of use are still rudimentary, requiring manual procedures so that the process can be executed optimally.

It is also worth noting that, in this work, we consider manual procedure as the act carried out by the indexer in the process of using controlled vocabulary accessing external environments to the platform in use (repositories, catalogs, and so on), that is, proprietary systems, spreadsheet documentation, and other types of digital resources. It is also worth noting that this research does not, at any time, deal with automatic indexing, and this study

is restricted to the integration of documentary languages to digital platforms of continuous use.

The main objective of this work is to present the use of SKOS ontology and of best publishing practices such as Linked Data as a computational artifact necessary to build controlled vocabularies for distributed use in decentralized environments. | 213

Also presented as specific objectives are the presentation of SKOS ontology, the basic concepts of Linked Data and also the suggestion of using controlled vocabularies by digital platforms.

SKOS (Simple Knowledge Organization System)

SKOS is a standardized model for publishing knowledge organization systems on the Web, based on the structural concepts of Semantic Web ontologies.

Although controlled vocabularies, and other types of documentary languages of this nature do not understand the use of axioms, inferences and instances, as do ontologies, SKOS model was conceived in such a way that technically, or technologically, a written documentary language using this standard is represented and mainly feasible for query and retrieval of data, such as the ontologies used in the Semantic Web standards.

It implies, therefore, that these structures of knowledge representation can be extended in its possibilities, with the use of relations that are used in the construction of ontologies.

There are a number of definitions for ontology concepts, some of them linked to philosophy or associated with the idea of taxonomy.

According to Santarem Segundo and Vidotti (2001, p. 287)

Ontologies provide structured knowledge and an infrastructure to integrate knowledge bases, independent of implementation. They are a powerful tool to support specification and implementation of computational systems of any complexity.

For the use as Semantic Web technology, ontologies are understood as: computational artifacts that describe a knowledge domain in a structured way, through: classes, properties, relations, constraints, axioms and instances.

SKOS has its specification, based on OWL language, as a complete ontology, which can be expressed using a knowledge representation model through the Resource Description Framework (RDF). Using RDF allows it to be transmitted across computer applications in an interoperable way. (SKOS, 2004).²¹⁴

According to W3C, RDF is a general-purpose language for representing information on the Web. RDF is intended to provide data interoperability so that it can contribute to information retrieval from resources on the Web.

The RDF model consists of three basic objects: resources, properties, and statements. A resource is information that can be identified by a Universal Resource Identifier (URI). Properties are the information that represents the characteristics of the resource, that is, the attributes that allow one to distinguish one resource from another or that describe the relationship between resources. The statement is the constitution of the complete information, which comprises a resource with its properties and values for the properties.

The SKOS model is divided into 9 groups: class concepts, schema concepts, labels, notations, properties, semantic relations, collections, and property mapping. These groups are represented by elements (such as: skos: prefLabel, skos: altLabel, among others) named vocabularies, which effect the relationship between terms in a controlled vocabulary.

Linked Data

Thinking about an associative model for publishing structured data on the Web, Linked Data was created (linked data). Organized by Tim Berners-Lee, the main characteristic is the establishment of links across data from distributed sources.

According to Heath and Bizer (2011), Linked Data is a set of best practices for publishing and connecting structured data on the Web, allowing links between items from different data sources to form a single global data space.

For Berners-Lee (2006),

The Semantic Web isn't just about putting data on the web. It is about making links, so that a person or machine can explore the web of data. With linked data, when you have some of it, you can find other, related, data.

The project has grown a lot in recent years, in 2007 Linked Data consisted of approximately one billion RDF statements, interconnected by 120,000 RDF links, consisting of approximately 25 datasets. In 2011 there were 52 billion RDF statements in 295 datasets. There are currently 570 datasets.

The largest and most prominent base in Linked Data is DBpedia. The main reference of instruments of vocabulary control and authority control in Linked Data is AGROVOC, which is a thesaurus developed from 1980, which currently is reference for subjects related to agriculture, pisciculture, forestry among other areas related to environment.

AGROVOC is used worldwide by researchers, librarians, information managers and other professionals to index, retrieve and organize data in agricultural information systems. Using SKOS terminology, an appropriate model for publishing thesauri on the web, AGROVOC can be consulted as well as interconnected with other resources to provide data in semantic format. (AGROVOC, 2014).

Using the SKOS model as well as Linked Data best practices, it is possible to build and make available any controlled vocabulary to the community, openly.

Construction and use of controlled vocabulary in a decentralized way

As observed, we currently have a set of technologies and tools that allow the construction and open publication of controlled vocabulary using SKOS. Using Linked Data best practices is one way to do this.

When analyzing the controlled vocabularies of most Brazilian institutions, evaluating the maintenance and mainly its use, we get to the problem that guides this research, the difficulty in making controlled vocabularies available in a decentralized way.

Some organizations use systems that allow the automated maintenance of controlled vocabulary, since there are, in the market automation systems of libraries, modules for this purpose. However, organizations that still use outdated tools or unusual techniques to build and maintain their controlled vocabularies are identified.

The use of controlled vocabularies by indexing teams on their digital platforms is even more complicated than the process of building and maintaining them. There are different processes and procedures for the use of controlled vocabularies in the various platforms that information units have to deal with daily.

In recent years, in addition to the well-known online catalogs (OPAC's), other tools and platforms have become part of organizations' daily lives and also require the use of controlled vocabularies for information organization. Scientific journals in electronic formats, digital repositories, integration platforms between libraries and publishers to access e-books, electronic document managers, among others, need the integration of controlled vocabularies for better organization and indexing of their content.

Information processing teams, still do not find, in these platforms, availability of controlled vocabulary for indexing the contents. In most cases, the procedure is performed manually and totally disintegrated. Most often, one chooses to access and query the controlled vocabulary in one system and manually insert the data into another system or platform.

This procedure can generate a great number of problems, from the most common ones, such as the incorrect typing of recurring terms to the difficulties with the periodical updating of terms, which generates an arduous task of updating the descriptors in the platforms that are not integrated in the documentary language.

Another problem is identified when there is a need to export controlled vocabulary built by one institution so that it can be used by another one. In these cases, there are difficulties with technologies that allow to load the structure of the controlled vocabulary. This difficulty is often caused because the documentary language was built in a proprietary tool or developed solely for a particular institution, and does not have a technical team responsible for updating or maintaining the tool.

When there is a possibility of exporting a vocabulary, it starts to have a life of its own in the institution that receives it, without contact with the creating institution, no longer receiving the updates and maintenance that the original continues to receive by the staff of the original institution.

Given the current context, the difficulties of using and maintaining controlled vocabularies and all the difficulties that lack integration, it is understood that there is a

latent need to organize a structure that allows the direct integration of controlled vocabularies to various types of platforms, regardless of the tool or system being used.

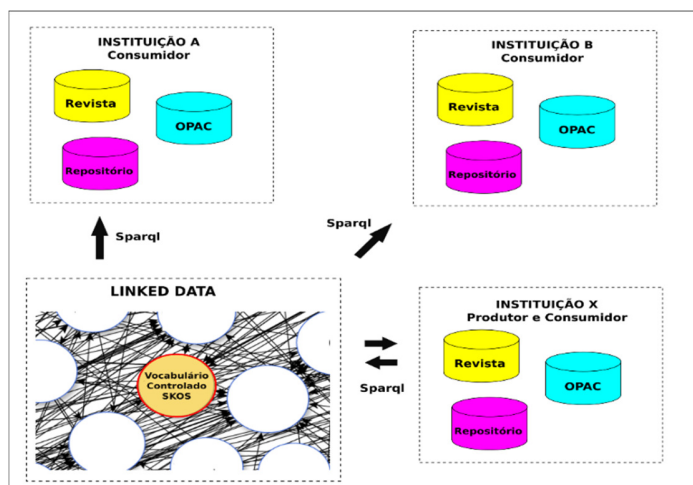
The proposal presented here is characterized by the use of a controlled vocabulary structure that is based on technologies that allow interoperability between systems and that controlled vocabulary is available in an open environment for free access by any and all digital systems or platforms.

It is understood that the use of Semantic Web technologies, based on RDF and organized by SKOS standard, allow the constitution of an interoperable and data-capable structure to be updated and retrieved by machines, through the integration of systems.

The proposal, as seen in Figure 1, assumes that one institution, or a group of institutions, called producers of the controlled vocabulary, perform the content maintenance tasks, while other institutions, named consumers, can integrate the product into all its platforms and tools for daily use.

The institution or group of institutions designated as producers will also be consumers of the controlled vocabulary.

Figure 1 - Decentralization model and integrated use of controlled vocabulary



Source: by the authors

Documentary language product, based on SKOS, will be published using the Linked Data best practices and will, therefore, be available for open and restriction free use with the application of an appropriate license.

In order to integrate the controlled vocabulary based on SKOS, it will be necessary to create plug-ins (pieces of code) that can be integrated to the most varied types of systems already mentioned, however, because the controlled vocabulary is built with open technologies, integration becomes fully possible.

Sparql protocol will be the main technology for data retrieving and updating by the producing and consuming institutions of the controlled vocabulary. Sparql is the name given to a set of specifications that provide languages and protocols to query and manipulate published content in RDF on the Web. Currently called protocol, version 1.1 is a W3C recommendation since March 2013. (SANTAREM SECOND, 2014) .

Conclusions

Based on the proposed objectives, this research presents a proposal for a construction model and use of controlled vocabularies based on Semantic Web technologies and concepts.

The use of RDF, Sparql and especially, SKOS model, is used as a propeller for the use of decentralized controlled vocabularies, and free of proprietary systems.

The solution based on the publication of controlled vocabulary using the best practices of open data, based on Linked Data, represents a new paradigm of construction and use of documentary languages by different systems and institutions, in a collaborative way. This new model has the principle of being decentralized and open, so that it can be integrated with various types of tools and platforms.

We conclude that the presented model, and the use of Semantic Web technologies, represent a new concept for the construction and use of controlled vocabularies.

References

The references was made following the ABNT rules.

AGROVOC. *AGROVOC Linked Open Data*. 2014. Available at: <<http://aims.fao.org/standards/agrovoc/linked-open-data>>.

BERNERS-LEE T.; LASSILA, O.; HENDLER, J. The semantic web. *Scientific American*, New York, v. 5, May 2001.

BERNERS-LEE, T. *Linked Data Principles*. 2006. Available at: <<http://www.w3.org/DesignIssues/LinkedData.html>>.

HEATH, T.; BIZER, C. *Linked Data: Evolving the Web into a Global Data Space*. Morgan & Claypool, 2011.

SANTAREM SEGUNDO, J. E. Web Semântica: introdução a recuperação de dados usando SPARQL. In: Encontro Nacional de Pesquisa em Ciência da Informação, 15, Minas Gerais, 2014, *Anais...* Belo Horizonte: Universidade Federal de Minas Gerais, 2014.

| 219

SANTAREM SEGUNDO, J. E.; VIDOTTI, S. A. B. G. Representação iterativa e folksonomia assistida para repositórios digitais. *Liinc em Revista*, v. 7, n. 1, p. 283-300, 2011. Available at: <<http://dx.doi.org/10.18225/liinc.v7i1.414>>.

SKOS. *Simple Knowledge Organization System*. Available at: <<http://www.w3.org/2004/02/skos/>>.

Preliminary analysis on the conversion of classification plans into controlled vocabularies



Luciana Davanzo
São Paulo State University
luciana.davanzo@gmail.com



Walter Moreira
São Paulo State University
walter.moreira@marilia.unesp.br

220

Introduction

Classification is one of the most important operations of archival work, as it underlies, in a way, almost all other operations. Two important archival tools are generated: a) the document classification plan, which aims to represent the information that was produced and received by the institutions in function of their activities, and b) the records schedule, which has, as objective, to establish the deadlines for keeping records.

The objective of the classification is basically to give visibility to the functions and activities of the organism that produces the archival, clarifying the links between documents, as Gonçalves points out (1998, p.12).

The relevance of the archival classification is observed from the publication by the National Archival Council (Conarq), linked to the National Archival, of "Classification, temporality and destination of archival documents related to the activities of the public administration", in 2001. This document is part of the process of "effective rationalization and control of information" (ARQUIVO NACIONAL, 2001, page 7) contained in public administration institutions.

Archival classification is one of the pillars of efficient and effective document management and allows, through the ordering of the relations between the subjects of the documents, greater precision both in relation to the process of information retrieval and in

relation to information control that are part of an institution. Thus, all types of documents, produced and received by the institutions, regardless of their media can and should be properly classified for later retrieval.

The effectiveness of information retrieval, therefore, is conditioned to quality in the processes of information representation, as widely known. Without clear and objective representation in relation to institutional interests, there can be no fully satisfactory retrieval by the user. ²²¹

Considering the value and importance of representation, both documentary representation performed by the professional and information representation performed by the user at the time of the search, it is believed that controlled vocabularies, in the archival classification plans, can act as effective mediators in the negotiations conducted by the system and user, mediated by language in archival information units.

In this view, we propose, as the objective for this research, to carry out an experiment of a controlled vocabulary "creation" from an established classification plan.

Although the classification plan acts positively in relation to access to documents, it is understood that for a more accurate information retrieval, it is necessary to make use of controlled vocabularies in conjunction with classification plans, as controlled vocabularies will guarantee that documents are described from terminological standardization, which will contribute to consistency in relation to their representation.

The development of the research is justified by the contribution of the use of controlled vocabularies by archivists as these are instruments that collaborate for the process of information organization and retrieval. Controlled vocabularies are still relatively little used by archivists, as they are still understood as auxiliary librarianship tools, an area they are most studied and applied.

In this way, it is believed that the instruments of control and terminological standardization add positively to document management process, since they stimulate care in relation to vocabulary control when naming, for example, the functions, activities and documentary series in the context of document management tools (SMIT; KOBASHI, 2003).

The analysis takes as a basis for the comparison and creation of controlled vocabulary, the subclass "020 Personnel Administration", of the classification plan

proposed for activities of the public administration by Conarq (ARQUIVO NACIONAL, 2001). This classification plan has ten main classes, which correspond to the great functions performed by the institution. Following the hierarchical decimal base, each class is sequentially subdivided into subclasses, groups, and subgroups.

Figure 1 shows a schematic representation of class 000 (General Administration) of the classification plan. Next, Figure 2 shows the first level of subclass 020 (Personnel Administration), which was used to compose the controlled vocabulary. The choice of this subclass was mainly due to the limits of the work presented here

Figure 1 - Organization of Class 000

CLASSE 000 – ADMINISTRAÇÃO GERAL
010 – ORGANIZAÇÃO E FUNCIONAMENTO
020 – PESSOAL
030 – MATERIAL
040 – PATRIMÔNIO
050 – ORÇAMENTO E FINANÇAS
060 – DOCUMENTAÇÃO E INFORMAÇÃO
070 – COMUNICAÇÕES
080 – (vaga)
090 – OUTROS ASSUNTOS REFERENTES À ADMINISTRAÇÃO GERAL

Source: Arquivo Nacional (2001, p.21)

Figure 2 - Organization of Subclass 020

020.1	LEGISLAÇÃO - Incluem-se normas, regulamentações, diretrizes, estatutos, regulamentos, procedimentos, estudos e/ou decisões de caráter geral e boletins administrativo, de pessoal e de serviço.
020.2	IDENTIFICAÇÃO FUNCIONAL (inclusive carteira, cartão, crachá, credencial e passaporte diplomático)
020.3	OBRIGAÇÕES TRABALHISTAS E ESTATUTÁRIAS. RELAÇÕES COM ÓRGÃOS NORMATIZADORES DA ADMINISTRAÇÃO PÚBLICA. LEI DOS 2/3. RAIS
020.31	RELAÇÕES COM OS CONSELHOS PROFISSIONAIS - Quanto aos documentos cujas informações gerem contenciosos administrativos ou judiciais, classificar no assunto específico.
020.4	SINDICATOS. ACORDOS. DISSÍDIOS - Quanto à contribuição sindical do servidor, classificar em 024.141. - Quanto à contribuição sindical do empregador, classificar em 024.153.
020.5	ASSENTAMENTOS INDIVIDUAIS. CADASTRO - Incluem-se documentos referentes à vida funcional do servidor, bem como os registros e/ou anotações a ele referidos. - Ordenar as pastas de assentamento individual alfabeticamente pelo nome do servidor. - Sugere-se abrir uma pasta para os documentos de caráter geral, sempre que tal procedimento se justificar.

Source: Arquivo Nacional (2001, p.23)

As a technical resource for this exploratory research, a software was created for the creation of thesauri called Thew, in Java version, created by Timothy C. Craven. Professor emeritus of the Faculty of Information and Media Studies, The University of Western Ontario. This software was chosen due to its ease of use and availability (free)

Controlled vocabularies and archival classification

In order to clarify classification and ordering, Heredia Herrera (1995: 263, free translation) defines classification as the action of "separating or dividing a set of elements by establishing classes or groups" and ordering as "uniting all the elements of each group following a unit of order, which can be by date, or the alphabet, or size or a number". Araújo (2006), discussing the commonality of classification, defines it as a process of "dividing into groups or classes, according to differences and similarities".

On this subject, Costa (1998 apud ARAÚJO, 2006, p. 118) observes that the action of classifying is a constitutive part of societies, as they are everywhere, impregnate social

life in a ubiquitous way, thus, the act of classifying accompanies the whole trajectory of man and is therefore necessary for development in relation to his worldview.

The activity *classification*, rather than merely a technical activity, must be understood in the broader scope of information organization and representation activities.

Knowledge organization and representation is a scientific-applied discipline

which aims to improve the circulation of information within and through mediating systems - the centers of selection, storage, retrieval, and dissemination of information - to produce new knowledge and/or to facilitate access to existing knowledge. (GARCÍA MARCO, 1993, p. 100, free translation).

One of the instruments used by information organization and representation to improve the information circulation is by controlled vocabularies. Controlled vocabularies allow the reduction of the use of various terms to represent one single piece of information or concept, as well as contribute to the reduction of communicational noises that may exist between users and information systems.

Controlled vocabularies are tools that cooperate both for the information organization process and the information retrieval process, so if used in classification plans and/or codes, they may be useful in relation to the polysemy of terms that may exist at the time of indexing and representing information.

Therefore, controlled vocabularies act as documentary language and are specially constructed to standardize and facilitate the entry and exit of data in an information system (Kobashi, 2008, p.1). Thus, it is believed that controlled vocabularies reduce the problems arising from information retrieval because they act directly in the control of terminological standardization.

Terminology control and standardization should be seen as opportunities for improvements in information retrieval, sharing, and socialization, since one of the main difficulties that archivists encounter is not the multiplicity of terms used to designate classification levels, but the lack of consensus on the meaning of these terms , as Sousa (2006, p.7) argues.

Especially in the case of non-specialized users, the use of controlled vocabulary becomes more than necessary, since these users will be able to carry out their research more proficiently, as well as avoiding information dispersion, which is a very common risk when there is no vocabulary control instrument.

For this reason, it is believed that classification is one of the great pillars of archival science. The instruments originated from it will serve as a foundation for other archival tasks, so it is understood that there is a need to use other instruments that contribute to the task information representation and retrieval, especially in times of technological advances, as classification cannot respond more to the important and updated task of searching for information contained in archival documents" (SOUSA, 2014, p.21).²²⁵

Results and discussion

The analysis developed in this section is based on the organization of subclass 020 of the Conarq Classification Plan (ARQUIVO NACIONAL, 2001), as shown in Figure 2. Based on this subclass and using Thew software, as described previously, the conversion of the terms treated in this subsection are converted into the structure of a controlled vocabulary. As far as possible, the same terms used in the Classification Plan were maintained, when necessary - to preserve the structure of the controlled vocabulary or for exemplification - new terms were inserted in the hierarchical, associative or equivalence relations.

Table 1 shows the result of the conversion in the systematic format of the controlled vocabulary. In this presentation mode, the relationship indicators between the terms (TG, TE, TR, USE, UP) and the scope notes (SN) are not shown. The choice for this presentation mode was based on the limits of the work set by the organization of the event.

Table 1 – Systematic presentation of controlled vocabulary

- ADMINISTRAÇÃO DE PESSOAL
 - ACORDO COLETIVO DE TRABALHO
 - ASSENTAMENTO INDIVIDUAL
 - CONSELHOS PROFISSIONAIS
 - CONSELHO FEDERAL DE ADMINISTRAÇÃO
 - CONSELHO FEDERAL DE BIBLIOTECONOMIA
 - CONSELHO FEDERAL DE MEDICINA
 - CONVENÇÃO COLETIVA DE TRABALHO
 - DISSÍDIO
 - IDENTIFICAÇÃO FUNCIONAL
 - CADASTRO FUNCIONAL
 - CARTÃO DE IDENTIFICAÇÃO
 - CARTEIRA DE TRABALHO E PREVIDÊNCIA SOCIAL
 - CRACHÁ
 - CREDENCIAL
 - PASSAPORTE DIPLOMÁTICO
 - LEGISLAÇÃO
 - BOLETIM ADMINISTRATIVO
 - BOLETIM ADMINISTRATIVO DE PESSOAL
 - BOLETIM ADMINISTRATIVO DE SERVIÇO
 - DECISÃO DE CARÁTER GERAL
 - DIRETRIZ
 - ESTATUTO
 - ESTUDO DE CARÁTER GERAL
 - NORMA
 - PROCEDIMENTO
 - REGULAMENTO
 - OBRIGAÇÕES TRABALHISTAS E ESTATUTÁRIAS
 - LEI DOS DOIS TERÇOS
 - RELAÇÃO ANUAL DE INFORMAÇÕES SOCIAIS
 - ÓRGÃOS NORMATIZADORES DA ADMINISTRAÇÃO PÚBLICA
 - SINDICATO
 - CONTRIBUIÇÃO SINDICAL
 - CONTRIBUIÇÃO SINDICAL DO EMPREGADOR
 - CONTRIBUIÇÃO SINDICAL DO SERVIDOR

Source: research data

In view of the network of conceptual relations established between the terms that were selected from the classification plan of the National Archival, it can be noted that the classification plan, from the use of controlled vocabulary, becomes clearer in relation to information representation.

Conclusion

Regardless of the institution and type of support, all documents must be classified because only the classification plan can give them the necessary sense of their use (LOPES, 1996, p.95). In this way, we highlight that, from the practice of vocabulary control, it will be possible to more effectively meet users' informational needs, because from the standardization of the terms used to represent the documents, it will be possible to create

a classification plan from a clear language, according to the terminology of the domain areas covered by Archival Science.

In future studies, we intend to extend the sample and extend the conversion experiment to other classes of the classification plan adopted as a model.

References

The references was made following the ABNT rules.

ARAÚJO, C.A.A. Fundamentos teóricos da classificação. *Encontros Bibli*, Florianópolis, n. 22, p. 117-140, jul./dez. 2006.

ARQUIVO NACIONAL (BRASIL). Conselho Nacional de Arquivos. *Classificação, temporalidade e destinação de documentos de arquivo relativos às atividades-meio da administração pública*. Rio de Janeiro: Arquivo Nacional, 2001.

GARCÍA MARCO, F. J. Paradigmas científicos en representación y recuperación de la información. In: GARCÍA-MARCO, Francisco Javier (Ed.). *Organización del conocimiento en sistemas de información y documentación*: actas del I Encuentro de ISKO-España, Madrid. Zaragoza: Universidad de Zaragoza, 1993.

GONÇALVES, J. *Como classificar e ordenar documentos de arquivo*. São Paulo: Arquivo do Estado, 1998.

HEREDIA HERRERA, A. *Archivística General: teoría y práctica*. 7. Ed. Sevilla: Servicio de Publicaciones de la Diputación de Sevilla, 1995.

KOBASHI, N.Y. *Vocabulário controlado: estrutura e utilização*. Brasília, Escola Nacional de Administração Pública, 2008.

LOPES, L.C. *A informação e os arquivos: teorias e práticas*. Niterói: Eduff. 1996.

SMIT, J. W; KOBASHI, N. Y. *Como elaborar vocabulário controlado para a aplicação em arquivos*. São Paulo: Arquivo do Estado; Imprensa Oficial, 2003.

SOUSA, R.T.B. Alguns apontamentos sobre a classificação de documentos de arquivo. *Brazilian Journal of Information Science: research trends*, v. 8, n. 1, 2014.

_____. Classificação de documentos arquivísticos: trajetória de um conceito. *Arquivística.net*, v. 2, n. 2, p. 120-142, ago./dez. 2006.

Thesaurus Reengineering: the Thesagro case



Benildes C. M. S. Maculan
Federal University of Minas Gerais
benildes@gmail.com

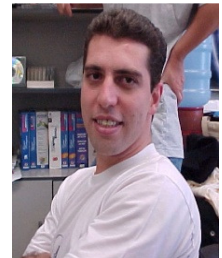


Gercina A. B. O. Lima
Federal University of Minas Gerais
limagercina@gmail.com

228



Ivo Pierozzi Jr.
Embrapa Agricultural Informatics
ivo.pierozzi@embrapa.br



Leandro H. M. Oliveira
Embrapa Agricultural Informatics
ivo.pierozzi@embrapa.br

Introduction

Information representation instruments for use in digital environments have been grouped under the name of Knowledge Organization Systems (KOS). In this environment, KOS "are now seen as schemes that organize, manage and retrieve information" (VICKERY, 2007, on-line) Thus, they have a single purpose: "to organize content to support the retrieval of relevant items, made available in the database of a digital library collection" (HODGE, 2000, p.9). To this end, it is important to print greater semantic richness in the conceptual structure of KOS.

Traditionally, the thesaurus is a type of KOS that has a semantic conceptual structure, because it is composed of a set of concepts that are linked together by different types of relationships. These relationships form a semantic network of relationships between terms and concepts, represented by equivalence relations (synonym terms and linguistic variations), identified by the symbols USE and UP (used for), hierarchical

(concepts ordered and grouped by different levels of generality and specificity), identified by the symbols BT (broader term) and NT (narrower term), and associative (other than equivalence or hierarchical relations), identified by the symbol RT (related term). The studies on the principles applied in the construction of thesauri have been evolving (MOTTA, 1987; CAMPOS, 1995; CAMPOS; GOMES, 2003, among others), and there are already initiatives that propose the specification of the different types of thesauri relations. ²²⁹

Among these initiatives, research on thesaurus conversion models in domain ontologies prevails, aiming to reuse the already structured knowledge of thesauri. In this context, this article reports part of the results of the application of the model developed by Dagobert Soergel et al. (2004) and Lauser et al. (2006) in the conversion of Thesaurus Agrícola Nacional (Thesagro) into a more formalized instrument, in such a way that the relationships of its conceptual structure are made explicit to the user. This thesaurus was chosen because it is the only Brazilian thesaurus in the area of Agriculture, since this study was carried out with the support of Embrapa Informática Agropecuária (Embrapa), a unit based in Campinas-SP, as part of a partnership signed between this institution, Federal University of Minas Gerais (Universidade Federal de Minas Gerais - UFMG) and the Hypertextual Map Prototype Research Group (Grupo de Pesquisa Protótipo Mapa Hipertextual - MHTX).

The thesaurus conversion model

The proposed model used in the conversion of Thesagro thesaurus has its procedures detailed in two articles: (1) Soergel et al. (2004) and (2) Lauser et al. (2006). This model was demonstrated by the authors in the conversion of a sample of the Agrovoc thesaurus and had, as objective, the creation of a heavy domain ontology, composed of classes, attributes and relationships between entities, all expressed in Web Ontology Language, Description Logic (OWL DL). As the authors' proposal was presented in a gradual way, it was possible to separate the conceptual phase of the thesaurus from the automation phase of the thesaurus conversion process into a domain ontology. Therefore, the results presented in this article report the intellectual procedures of refinement and explication of the existing relations in the conceptual structure of Thesagro thesaurus.

The main characteristic of this thesaurus-reengineering model is to allow the individualization of the modeling in five levels of entities: concept, term or lexicalization, string or variants, scope notes and relationships. In this way, each level indicates different types of information and the relationships, considered the backbone of the thesaurus, can be assigned among entities of the same level (for example, among different concepts) or among entities of different levels (for example, among terms and strings), as follows:

concept to concept	is_a (hierarchy); plague_of
term to term	is_synonymouse; is_tranlsation_of
concept to terms	has_lexicalization (links the concepts to its lexical representation)
Term to string	has_acronym; has_variation_orthographic; has_abbreviation (links the terms to its variant forms)

The model consists of three basic steps: (1) definition of the structure of the converted thesaurus, using an existing thesaurus in the domain to be worked; (2) collection of terminology and other information, from one or more thesauri, in the domain to be modeled; (3) editing the thesaurus, with the reformulation of the existing thesaurus, transforming its structure into a more semantic conceptual network.

Application of Thesaurus Conversion Model Steps

For the application of the thesaurus conversion model, it was necessary to make an initial planning in which a work team was constituted by five members: a modeler (information professional), a domain expert, two terminologists and a technologist.

The first step, from the definition of the structure of the converted thesaurus, covered the mapping of Thesagro characteristics. We found that this thesaurus has about 9,400 descriptor terms, preferred and non-preferred, and it is considered a high specificity and broad scope thesaurus. Similar to other traditional thesauri, Thesagro presents the three basic relationships using symbols in English language:

Equivalence Relations:	USE and USED FOR (UF)
------------------------	-----------------------

Hierarchical Relations:	BROADER TERM (BT) and NARROWER TERM (NT)
Associative Relations:	RELATED TERM (RT)

By the Thesagro thesaurus analysis, we observed the presence of about 2,000 preferred descriptors, which did not belong to any cluster in the Thesagro hierarchical structure, being considered orphan descriptors. This lack of linkage between concepts can sometimes make the semantics of the thesaurus structure less understandable to the user.

The first stage also included the determination of the subarea of Agricultural Intensification as the thematic clipping for the modeling. The conceptualization of this theme follows Boserup's (1965) theory, according to which there is a relation between the dynamics of the population of a region, with its growth or not, the environment and the use of technology in agricultural production, originating the growth or maintenance of a production at a given time and place.

The modeling of this subarea was based initially on a taxonomy elaborated by Embrapa experts, composed of about 600 concepts and was structured in nine categories: (1) extensive agriculture; (2) intensive agriculture; (3) material and methods; (4) environment; (5) agronomy; (6) territory and landscape; (7) socioeconomics; (8) space and time; (9) institutions. Among the concepts of the taxonomy, a sample composed of 30 concepts representative of each of the nine categories was defined and, thus, of the subarea of Agricultural Intensification as a whole.

In the second stage, from the terminology collection, the taxonomy and three thesauri were used as terminological inputs with agricultural scope: Thesagro, Agrovoc and National Agricultural Library (NAL). For this activity, a comparison was made among the concepts of the sample and the existing terminology in the taxonomy and in each of the three selected thesauri. This was performed from two lists: (1) list composed by the 30 concepts of the sample, in Brazilian Portuguese, adding the expressions of these terms in the singular and plural; (2) list composed by the 30 concepts of the sample, translated into English, adding the expressions of the terms in the singular and plural, and also in its inverse form (adjective + noun), because this inversion is common in English. The terminological comparison was carried out in a semiautomatic way, based on an

intellectual analysis of the instruments and an automatic processing of comparison. This processing was carried out using the Multilingual Agricultural Conceptual Terms and Structures Extractor (Extrator de Termos e Estruturas Conceituais Agrícolas Multilíngue - ETECAM), which allowed the retrieval of the terms that coincided with the two lists in the instruments, as well as their semantic clusters, separately. | 232

The third stage involved the editing of the remodeling of Thesagro's conceptual structure, using the Electronic Terms software (e-Terms), which is a tool that supports the creation and management of terminological products for different purposes (teaching, glossaries, controlled vocabularies). The e-Terms system was developed as a collaborative web-computing environment, which can be used free of charge, but restricted to registered users. It is composed of six stages, with automated and semi-automated procedures, based on the principles of the Communicative Terminology Theory (TCT), developed by Cabré (1999). TCT has a theoretical construct that defends a similarity between the systems of the specialty language and of the general languages, since both are governed by the same rules and are characterized by the same phenomena of synonymy and linguistic variation. These principles minimize the normative rigidity derived from the General Theory of Terminology (TGT) developed by Wüster (1998), whose foundations are generally used in the construction of thesauri.

The e-Terms system allowed the creation of a Definitional Base that compiled and stored explanatory excerpts and/or the concepts of the sample. The content of these excerpts helped to make a glossary with the definitions of the sample concepts. In addition, the e-Terms also allowed the elaboration of terminological data sheets for these concepts, which was composed of 38 semantic fields, among which we highlight the fields for definitions of concepts (specialist, modeler and final), for encyclopedic information and of gloss, for scope notes and for terms in relation to equivalence, linguistic variation, as well as fields for concepts in hierarchical and associative relationship.

In the construction of the concept system for the concepts of the sample and their semantic clusters, 44 different types of relations were used. This generated an unfolding that totaled the representation of about 600 relationships: of gender and their species, of the whole and its parts, of equivalences, of strings (variations) and associative.

For an example of the current structure of Thesagro and proposals for reformulation using the thesaurus conversion model, the current modeling of the descriptor *GATO* (cat in English) in Thesagro follows:

GATO <i>BT</i> MAMÍFERO DOMÉSTICO <i>NT</i> GATO ANGORÁ <i>NT</i> GATO DO MATO <i>RT</i> FELIS CATTUS DOMESTICUS <i>RT</i> FELIS DOMESTICA	FELIS CATTUS DOMESTICUS <i>RT</i> GATO
	FELIS DOMESTICA <i>RT</i> GATO

In the current structure of Thesagro, it is noted that, for example, the descriptor *MAMÍFERO DOMÉSTICO* is a more general concept and that the descriptors *GATO ANGORÁ* and *GATO DO MATO* are more specific concepts of *GATO*. Regarding the other relationships, it is possible to notice that the preferred descriptor *GATO* is in an associative relationship with the preferred descriptors *FELIS CATTUS DOMESTICUS* and *FELIS DOMESTICA*, without the specific type of associative relationship being clear.

Applying the thesaurus conversion model, the reformatting of the model of the descriptor *GATO* has the following configuration:

GATO <i>temTermoGenérico</i> MAMÍFERO DOMÉSTICO <i>temTermoEspecífico</i> GATO ANGORÁ <i>temTermoEspecífico</i> GATO DO MATO <i>temNomeCientífico</i> FELIS CATTUS DOMESTICUS <i>temNomeCientífico</i> FELIS DOMESTICA	FELIS CATTUS DOMESTICUS <i>temNomePopular</i> GATO
	FELIS DOMESTICA <i>temNomePopular</i> GATO

With this reformulation, the type of relationship that was established between the concepts became explicit, facilitating the understanding of the semantic structure. Thus, any user can apprehend that the descriptor *MAMÍFERO DOMÉSTICO* is a more generic term and is in a relation of genre-species with *GATO* that, in turn, is in a relationship also of genre-species with *GATO ANGORÁ* and *GATO DO MATO*. In addition, we now clearly notice that the descriptors *FELIS CATTUS DOMESTICUS* and *FELIS DOMESTICA* are scientific nomenclatures of the descriptor *GATO*, which is a popular name of this animal.

In this way, we can affirm that the current Thesagro model loses the semantic function that connects the descriptor GATO with the descriptors that represent the scientific nomenclatures. This type of structure is repeated in the modeling of all living organisms included in the terminology of the current Thesagro and explains the existence of about 2,000 orphan descriptors. With this form of modeling, Thesagro's current representation keeps dispersed the semantic sense of its conceptual structure, making it difficult to understand the knowledge domain, unless the user already possesses such knowledge *a priori*.²³⁴

With the reformulation of the concepts of the sample and its semantic clusters, 286 occurrences of hierarchical relations were represented, being 225 genre-species relations (with 52 general terms and 173 specific terms) and 61 whole-part relations (with 22 general terms and 39 specific terms), which shows that the new structure of Thesagro maintains its characteristic of thematic specificity. In addition, 232 associative relationships were represented, confirming that the subarea of Agricultural Intensification is complex in nature, with many interrelated concepts.

The results evidenced that the explicit expression of the relations among entity pairs (concepts, terms, strings and scope notes) refined the semantics of the structure of the converted thesaurus, giving subsidies to facilitate interoperability among different thesauri or systems. We emphasize that the complete results of the research can be consulted in the thesis of the leading author of this article.

Conclusions

The development of this study had, as its starting point, the semantic enrichment of the conceptual structure of a traditional thesaurus, aiming to make it a type of KOS, denomination that covers the instruments that can be understood by the machine.

During the course of research, it was evident that the construction or reformulation of a thesaurus is quite complex, which requires the involvement of different professionals, such as librarians, terminologists and specialists. In addition, we verify the importance of meeting the needs of the user, since the terminology of a thesaurus, even if it is a specialized language, should prioritize the user's search language.

We also realize that the development of the semantic structure of a thesaurus must be guided by international norms and standards for the construction of thesauri, since they establish principles for a more formal representation, which can aid in the interoperability among different vocabularies and systems.

We consider, as the main contribution of this research, the validation of a traditional thesaurus conversion model, which can be applied as a data integration solution in different systems. We evaluate that this can positively impact the area of Information Organization, advancing the studies about methodologies that can be used in the construction of thesauri.

References

The references was made following the ABNT rules.

BOSERUP, E. *The conditions of agricultural growth: the economics of agrarian change under population pressure*. Chicago: Aldine, 1965.

CABRÉ, M. T. *La terminología: representación y comunicación*. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra, 1999.

CAMPOS, M. L. A. Linguagens documentárias: núcleo básico de conhecimento para seu estudo. *R. Esc. Biblioteconomia UFMG*, Belo Horizonte, v. 24, n. 1, p. 52-62, jan./jun. 1995.

CAMPOS, M. L. A.; GOMES, H. E. Organização de domínios de conhecimento e os princípios ranganathianos. *Perspect. Ci. Inf.*, Belo Horizonte, v. 8, n. 2, jul./dez. 2003.

DAHLBERG, I. Teoria do conceito. Tradução Astério Tavares Campos. *Ci. Inf.*, Rio de Janeiro, v. 7, n. 2, p. 101-107, 1978.

HODGE, G. *Systems of knowledge organization for digital libraries: beyond traditional authorities files*. Washington, DC: Council on Library and Information Resources, 2000. Available at: <<https://www.clir.org/pubs/reports/pub91/pub91.pdf>>.

LAUSER, B. et al. From AGROVOC to the Agricultural Ontology Service: Concept Server an OWL model for creating ontologies in the agricultural domain. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 2006, Colima, Mexico. *Proceedings...* México: DCMI, 2006.

MOTTA, D. F. da. *Método relacional como nova abordagem para a construção de tesouros*. 1987. 89f. Dissertação (Mestrado em Ciência da Informação) – Instituto Brasileiro de Informação em Ciência e Tecnologia, Rio de Janeiro, 1987.

SOERGEL, D. et al. Reengineering thesauri for new applications: the AGROVOC example. *Journal of Digital Information*, v. 4, n. 4, 2004.

VICKERY, B. C. A note on knowledge organization. [2007]. Site Lifeboat for Knowledge Organization. Available at:
<http://www.iva.dk/bh/lifeboat_ko/concepts/Vickery_a_note_on_knowledge_organisation.htm>.

WÜSTER, E. *Introducción a la teoría general de la terminología y a la lexicografía terminológica*. Barcelona: IULA, 1998.

Conceptual Model of Organizational Competitive Intelligence



Thiciane Mary Carvalho Teixeira
Ceará State University
thiciane.teixeira@uece.br



Marta Lgia Pomim Valentim
So Paulo State University
valentim@valentim.pro.br

237

Introduction

Organizational strategies such as the systematic search for external and internal information that enables managers to get to know their own company, the market, the competition and the customers, and the analysis of data in order to select, share and disseminate tangible (explicit) knowledge and, from its appropriation, to construct intangible [tacit] knowledge, has become differential in competitive business environments. Knowledge, information and information and communication technologies (ICT) have become significant assets for generating organizational competitive intelligence (OCI).

From the collection and analysis of data and information, it is possible to filter and treat relevant content to the organization, whose added value from the analysis is directed to decision making and planning of strategic actions. It is possible to transform information and knowledge into intelligence, whose goal is to provide continuous improvement of organizational performance, enabling problem solving, creating new products, designing change and innovation.

OCI refers to a strategic process that focuses on the conversion of data and information on capabilities, vulnerabilities and intentions of the competition into a competitive advantage for the organization. It is a process that needs to be constantly built, maintained and evaluated, as it is driven by a dynamic that involves information

search, sharing and use in a creative way, aiming to discover opportunities, reduce risks and provide different types of combined innovations involving quality, productivity and competitiveness.

Level of responsibility and behaviors applied to OCI

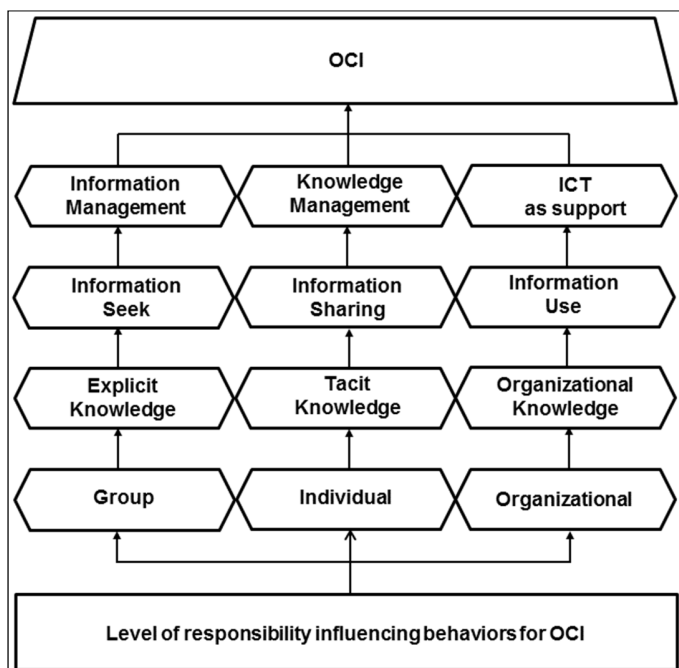
The proposal highlights the identification of three different levels of responsibilities that influence the behavior of search, sharing and use of information and ICT for OCI:

a) **Individual**: through the valuation of individual knowledge, skills and competences. It refers to the set of tacit knowledge of managers, individually, coming from formal education, experiences and entrepreneurship. A relationship of positive influence of this level of responsibility with the behavior of information and ICT search, sharing and use was evidenced.

b) **Group**: through the socialization of information, it is possible to create a group understanding about a given information or problem. It refers to the knowledge of different organizational subjects made explicit and available to the group of managers. A relationship of positive influence of this level of responsibility with the behavior of information and ICT search, sharing and use, and with the subtleties of the group in relating, organizing and treating the available information was evidenced.

c) **Organizational**: by understanding the impact of individual or group actions, as well as responsibility for the company's results and competitiveness. It refers to organizational knowledge, including the influence of organizational culture and information in practices of searching, sharing and using of information and ICT. A relationship of positive influence of this level of responsibility with the behavior of information and ICT search, sharing and use as a support for planning and decision making was evidenced.

Figure 1: Level of responsibility influencing behaviors for OCI

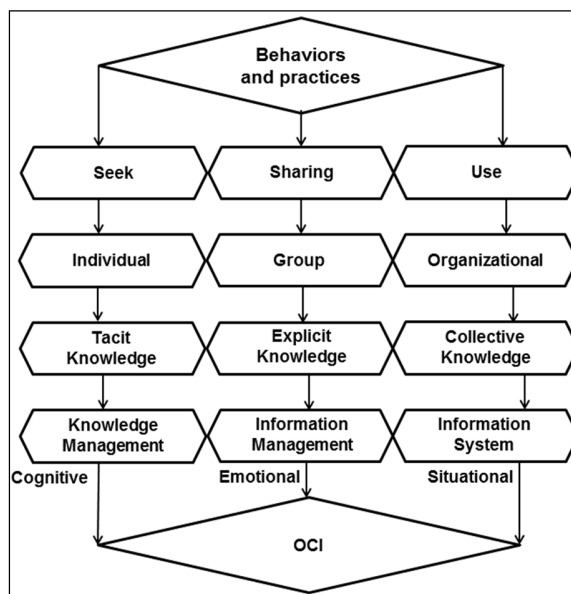


Source: By the authors – 2014

The level of individual responsibility corresponds to human capital, it refers to the tacit knowledge of the organizational subject that can be externalized and managed in the organizational context. The manager, as a leader and entrepreneur, represents this level, that is, it is through the interpretations of facts and events, and information sharing, it is possible to generate new knowledge that can be appropriated by other organizational subjects with meanings and understandings proper to experienced organizational reality, which makes it possible to solve problems or to prepare to anticipate external and internal opportunities and threats to the company.

The level of responsibility of the group corresponds to the construction of meaning and understandings, through the collective understanding of information to be managed in the organizational context. Sources of information play a vital role, as suppliers and associative networks participate, generate, and deliver business-relevant information.

The level of organizational responsibility corresponds to the physical and technological structure that allows information access and use, and most importantly, the organizational and informational culture, which is essential to guarantee OCI practice in the company.

Figure 2: OCI behaviors and practices

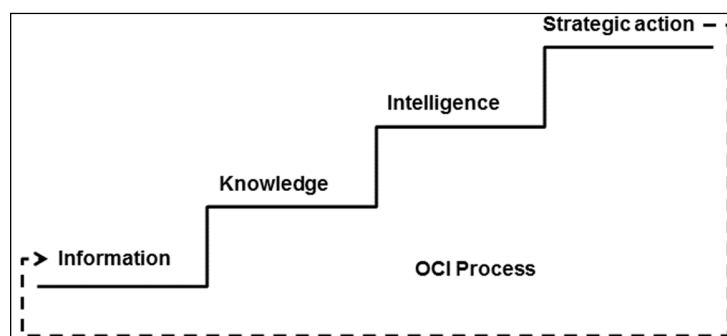
Source: by the authors – 2014

OCI is a process of orientation toward strategic action, be it for the allocation of resources, for the planning of future actions, to provide innovation in products, services and processes. The organization needs to develop information competencies in order to manage the relevant information to the business and also the knowledge generated by the organizational subjects of all hierarchical levels in order to obtain competitive differentials.

OCI process

OCI is understood as a learning process, since information needs to be analyzed and knowledge must be appropriated by the organizational subjects, in order to make inferences regarding the reality experienced by the company. OCI is a continuous process that, after appropriation of the relevant information and knowledge internalization by the group, establishes a dynamic of capturing, adapting, creating, innovating, perceiving and acting in a strategic way.

Figure 3: OCI process



Source: by the authors – 2014

OCI contributes to competitiveness as it is an ongoing process, based on behaviors and practices in information and ICT search, share and use as consistent and systemic learning, built on the belief that it is possible to innovate, learn and act intelligently, even if there is a risk at distinct degrees regarding the decision to be made.

OCI process establishes three guidelines that result from behaviors applied to OCI (Figure 4):

1) *Guidelines for information management*

It is based on phases 1 and 2 defined by Miller's model (2002), which correspond to the identification of the needs and the actions of selecting and filtering relevant information to the business, in external and internal environments. This guideline positively influences the level of individual responsibility for information-seeking behavior. The importance of motivation and training to sensitize and recognize explicit knowledge present in the information available in the organizational context, focusing on information management is highlighted.

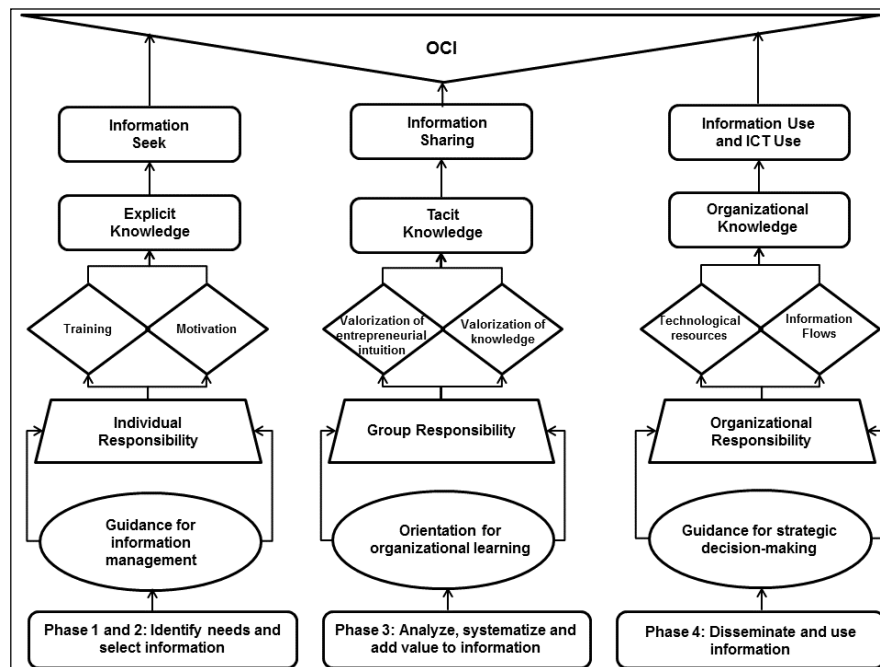
2) *Guidelines for organizational learning*

It is based on phase 3 defined by Miller's model (2002), which corresponds to the steps of analysis, systematization and adding value to information. This orientation positively influences the group's level of responsibility for information sharing behavior. We highlight the importance of valorizing tacit knowledge present in the intuitions and entrepreneurial characteristics of each organizational subject and the group for accomplishing information sharing as a foundation that fosters organizational learning.

3) *Guidelines for strategic decision making*

It is based on phase 4 defined by Miller's model (2002), which corresponds to the stage of information dissemination and use to generate organizational intelligence. This guideline positively influences the organization's level of responsibility for information and ICT use behavior for strategic decision making. It is important to highlight the importance of the use of technological resources as support, and information flows and communication channels as structures that stimulate organizational and informational culture to guide decision making.

Figure 4: Conceptual model of OCI



Source: by the authors – 2014

The proposed conceptual model of OCI can be applied from the answers to three questions related to the behaviors and practices of information and ICT search, share and use:

- *How does the information reach the company?* Organizational subjects need to assess information sources as to credibility for selecting and filtering relevant content. The obtained information must be reliable and consistent in order to serve as input to the decision-making process, so managers need to believe that using the available information will produce more effective result

- *How is information shared for strategic management?* New information appropriate for the group should be added to the previously existing set of knowledge. Individuals' intuition and perception from their experiences can be united to this information, in order to provide conditions to evaluate and to decide if information can be useful or not for strategic action. | 243
- *How can technological resources be applied to decision making?* The use of technological resources is important for the search, sharing and use of information in the organizational context, since they provide agility and dynamics to the decision-making process.

Material and methods

Qualitative, descriptive and explanatory research applied a structured questionnaire and interview script, allowing the description and explanation of the phenomenon in an attempt to acknowledge and expose the performed practices and the importance of the behaviors for OCI process.

The deductive method was based on the following hypotheses: a) information search and sharing behaviors positively influence prospecting and monitoring activities; b) the use of information and ICT positively influence the activities of decision-making and planning of strategic actions.

The universe surveyed included 29 (twenty-nine) companies from Ceará awarded in 2012 with Delmiro Gouveia Prize, which establishes that the largest companies should be those with better economic-financial and social performance in the year, evaluated according to the net revenue: over R\$ 90 million and below than R\$ 90 million.

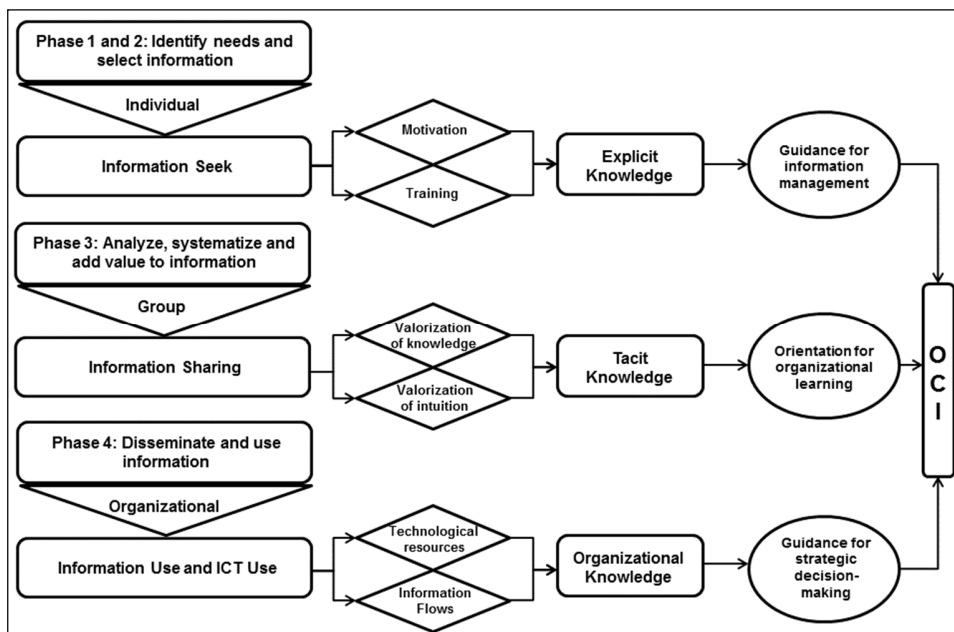
For data analysis, Content Analysis by Bardin (1977) and simple statistical analysis processed by Excell software was applied. The procedure allowed the presentation of indicators through graphs and tables.

Results

The conceptual model of OCI (Figure 5) can serve as a guideline for companies operating in competitive markets. It can also be realistically adapted to smaller companies.

Organizational Competitive Intelligence that one wishes to achieve lies in the fact that the behaviors of information and ICT searching, sharing, using are determining practices for obtaining competitive advantage.

Figure 5 - Conceptual model of OCI



Source: by the authors – 2015

OCI emerges as the result of behaviors and practices, levels of responsibility, and guidelines that can serve as model. It should be emphasized that organizational and informational culture is fundamental for OCI process, since organizational subjects need to value the search, sharing and use of information and ICT to perceive competitive advantages.

The hypotheses that information search and sharing behaviors influence prospecting and monitoring activities, and that the use of information and ICT influence decision-making activities and strategic action planning were positive.

Companies know what is competitive intelligence, routinely use and practice OCI phases to obtain competitive differentials. There are some difficulties for companies to maintain and systematize the practices of information and ICT search and sharing that involve cultural variables, which emerge as an important factor to be worked towards the application of competitive intelligence.

Information retrieval practices are considered survival strategies. Searching information with clients is one of the main pursued practices, aiming to guarantee the quality services and increase competitiveness. It was evidenced that information search with customers is more constant than it is with the competitors.

Concerning for internal information emerges as a more systematic practice of information searching. There is concern about the control of internal information, as it directly influences the tactical and operational level.

Monitoring activities are a reality and search forms are varied. Political and economic monitoring proved to be an essential action. It has been shown that companies perceive government more as a threat than market competition.

The entrepreneurial attitude of managers is an important variable for systematizing and adding value to information, contributing to real and effective use of information. After information search by the manager, information sharing is carried out and influences the organizational group subsidizing decisions and planning. In addition to the individual information search at the strategic level, the effective use of ICT in internal networks and organizational channels was also identified, through information flows to improve processes, activities and daily tasks.

The companies make use of information and communication systems, aiming at information access, storage and dissemination to support decision making and planning, as well as the establishment of strategic actions.

Considerations

The elaborated conceptual model of OCI deserves to be validated to be recognized theoretically. The application of the model and evaluation of the practices of information and ICT search, sharing and use can have an impact on the financial results of the company.

The study proposes actions through motivation, valuing knowledge and information flows as structures necessary for OCI process. Behaviors of information and ICT search, sharing and use lead to entrepreneurial practices and the intelligent application of information and knowledge in strategic format, providing an impact on the organization results with a competitive advantage. | 246

The application of the OCI model will allow the optimization of strategic decisions, organizational learning and information management, with the involvement and participation of the individual, the group and the organization.

References

The references was made following the ABNT rules.

BARDIN, L. *Análise de conteúdo*. Lisboa: Edições 70, 1977. 226p.

MILLER, J. P. *O milênio da inteligência competitiva*. Porto Alegre: Bookman, 2002. 293p

TEIXEIRA, T. M. C. *Inteligência competitiva organizacional: um modelo apoiado nos comportamentos de busca, compartilhamento e uso de informação e de TIC*. Marília, 2014. 200f. Tese (Doutorado) – Programa de Pós-Graduação em Ciência da Informação – Universidade Estadual Paulista (Unesp), 2014.

Vocabulary control in electronic scientific journals



José Carlos Francisco dos Santos
Londrina State University
jc.fa1982@gmail.com



Brígida Maria Nogueira Cervantes
Londrina State University
brigidacervantes@gmail.com

247

Introduction

In the current context of technology, specifically with information on the web, journals are already created in the electronic medium. In this environment, the process of choosing keywords for representing thematic content deserves special attention as they are essential resources for access and retrieval of content in electronic scientific journals.

The aim of this study is analyze the operating conditions of a vocabulary control of keywords assigned in electronic scientific journals through VCI&I tools, as a complementary tool for IBICT's Electronic Journaling System (SEER). This proposition is justified by the effects produced to improve the retrieval of contents indexed in electronic scientific journals, available in SEER system, in addition, serving as a consultation instrument so authors can choose the most relevant and consistent terms for disseminating their scientific research productions.

Vocabulary control in electronic scientific journals

Keywords play an important role in thematic content retrieval in scientific publications. The submission rules commonly include guidelines for authors, which may

be used as basis to design the keywords as best as possible, an indication for a reliable representation of their text (GONÇALVES, 2008).

In the international standard ANSI/NISO Z39:19-2005, the keyword is defined as the expression that appears in the text, configured through natural language, and is considered the most significant for indexing or retrieving the text. (AMERICAN NATIONAL STANDARDS INSTITUTE; NATIONAL INFORMATION STANDARDS ORGANIZATION, 2005). According to NBR 6028, the keyword is the word representative of the document content, preferably chosen in controlled vocabulary (BRAZILIAN ASSOCIATION OF TECHNICAL STANDARDS, 2003b, p.1). For Miguéis and Neves (2013), keywords distance themselves from the indexing term, because the latter has the result centered in the process of conceptual analysis that translates the controlled vocabulary and, in turn, the keywords assignment is free, being chosen by the author, editor or even generated automatically.

Gil Leiva and Alonso Arroyo (2005, our translation) state that the product of indexing professionals should not be replaced by keywords assigned by authors, but keywords are an important source for the analysis of scientific articles in indexing processes, as well as an important source for beginning the methodologies of automatic indexing.

Vocabulary control [...] has generally been exercised through: selection of preferred terms; distinction of homographs; formal establishment of relations among terms. (FUJITA, 2011, p.42, our translation). Vocabulary control is used to increase the effectiveness of information storage, retrieval systems, Web browsing systems, and other environments that need both to identify and to locate the desired content of some type of description using language. The main purpose of vocabulary control is to achieve consistency in the description of content objects and facilitate retrieval (AMERICAN NATIONAL STANDARDS INSTITUTE; NATIONAL INFORMATION STANDARDS ORGANIZATION, 2005).

The need for vocabulary control is due to two basic characteristics of natural language: 1) duplicity of nomenclature for the same concept; 2) duplicity of concept for the same word. To obtain vocabulary control, three main methods are necessary: a) to

define the scope of the terms; b) to use the equivalent relationship for the connection of terms with synonyms and almost synonyms; and c) to distinguish between homographs (AMERICAN NATIONAL STANDARDS INSTITUTE; NATIONAL INFORMATION STANDARDS ORGANIZATION, 2005).

Methodological procedures

VCI&I tools was developed in PHP language, it is a free language for web and it performs perfect integration with MySQL database management system. The tool performs the following procedures, here identified as of functional requirements, listed with the Acronym RF: RF01 Metadata collection; RF02 Metadata Analysis and Controlled Vocabulary; RF03 Keyword Treatment; and RF04 Results.

It should be clarified that, in order to prepare VCI&I tools, we chose to use consolidated controlled vocabularies, available electronically, which would be required in the metadata analysis functionality (specifically the keywords) with controlled vocabulary. In choosing the controlled vocabularies for the confirmation of terms, it was considered, as the main parameter, publication actuality and consistency, being those offered by Pinheiro and Ferrez (2014) - *TBCI* and Moreira (2006) - *TCI*.

The main purpose of RF01 is to collect the selected journal metadata through OAI-PMH protocol and store it in the database. The purpose of RF02 is to check the keywords that refer to the same terms according to controlled vocabularies, using the following comparison mechanisms:

I) identical - performs the comparison between the keyword and the controlled vocabulary term;

II) identical, removing empty words - empty words are excluded from the keywords and controlled vocabulary terms, for example: definite and indefinite articles, prepositions, among others.

III) identical removing plurals - plurals are treated, using the set of rules and exceptions by Orengo; Buriol and Coelho (2007), who propose improvements to the stem-translation algorithm in Portuguese RSPL by Orengo and Huyck (2001) - Suffix Remover.

IV) identical removing suffixes - depends on the reduction of empty words and plural ones, the rules and exceptions are by Orengo; Buriol and Coelho (2007).

V) identical by contained index - from the contained index to be calculated and stored in the database between each keyword and each controlled vocabulary term.

The purpose of RF03 is to perform the treatment of keywords which are not compatible with controlled vocabularies with human intervention to manage this process. The objective of RF04 is to show quantitative results of the processes related to keywords compatibility with controlled vocabularies. ^{1,250}

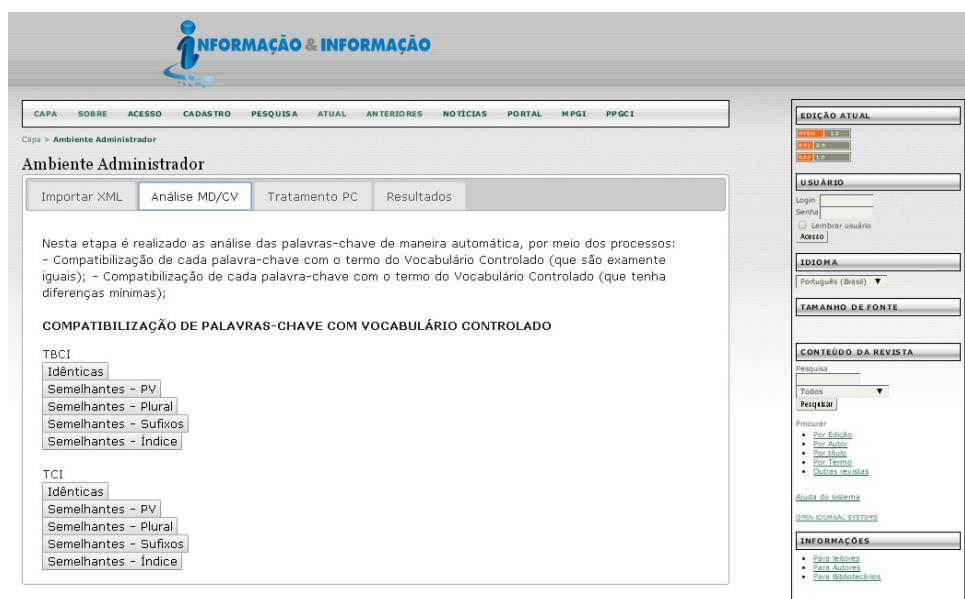
The choice for the application of the VCI & I tools prototype was the Journal *Informação & Informação* for being a scientific journal of Information Science and is assessed by Qualis, currently A2. *Informação & Informação* is linked to the Graduate Program in Information Science, Information Science Department, Londrina State University.

Data were collected in the metadata of published articles, from the first issue of 1996 to the last issue of 2014. A scientific journal includes several types of sections, such as articles, book/media reviews, theses and dissertations, points of view, experience reports, communications. Therefore, the expression *articles* was assigned as a generic designation for the sections.

Application of the proposal: analysis and results

In order to achieve the indicated results, the processes of VCI & I tools were executed to import metadata. In all, the import consisted of 437 records (documents), totalizing of 1199 found keywords, from which a total of 365 identical keywords were identified, and 834 keywords were imported into the VCI&I database tools that were compared in the compatibilization process. Figure 1 shows the analysis interface of the keywords.

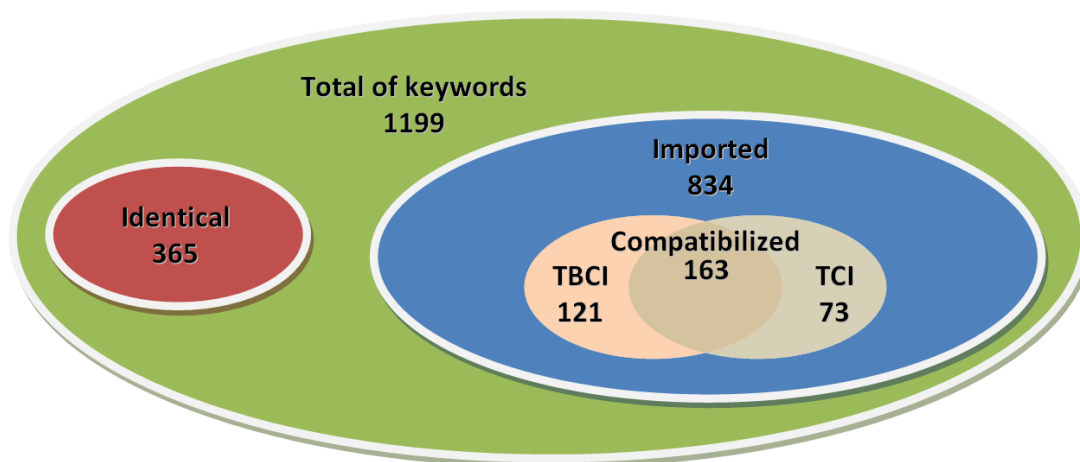
Figure 1- metadata analysis interface - keywords



Source: by the authors, using SEER system layout

The compatibilization process was performed through the interface of VCI&I tools described in RF02. Therefore, it can be observed a total compatibilization of 121 keywords were found with *TBCI*, and 73 keywords with *TCI*. The total of matched words was 357, shown in Figure 2. Unmatched keywords were treated manually. From the automatic compatibilization and manual treatment of the keywords, the index of terms is available through VCI & I tools.

Figure 2 - Results from import and keywords compatibility processes.



Source: by the authors

In this index shown in Figure 3, the user has access to the preferred terms arranged in alphabetical order, grouped by the initial letter of the terms and also by localities and personal names. In each term, there are the number of articles related to it. Both the term and the quantity of related articles are links, which by clicking on the term, it is presented its related keywords and the appropriate compatibilizations with *TBCI* and *TCI*, if applicable. By clicking on the number of articles related to this term, the user has the title of each linked article (through a link) to the article in case one wants to consult it.

Figure 3 - List of terms interface.



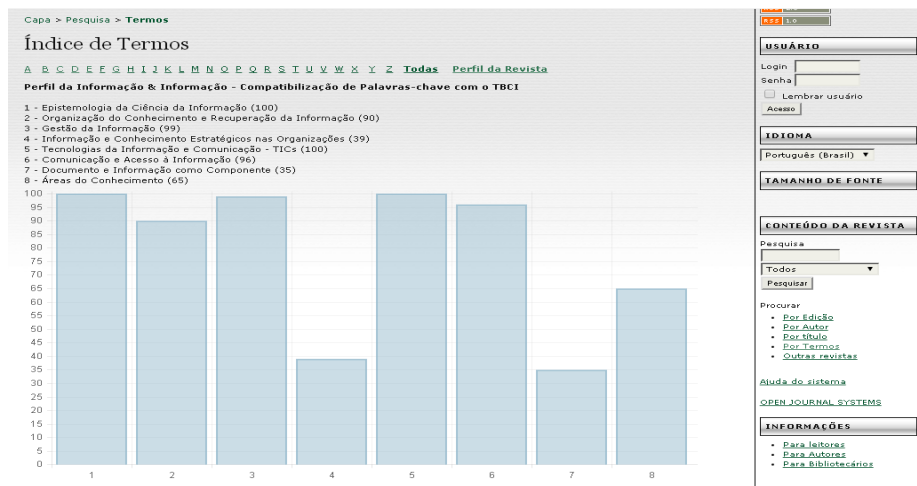
Source: by the authors, using SEEL system layout

In the user environment, it is possible to access the profile of the journal traced through the main categories that fit the keywords of *TBCI*'s General Classification Plan. The second profile is formed through *TCI*'s facets. It is observed that in both profiles, a list of categories/facets and the number of related articles are presented through keywords compatibility.

Data from the profile of the journal, according to *TBCI* categories are: 1 - Epistemology of Information Science (100); 2 - Knowledge Organization and Information Retrieval (90); 3 - Information Management (99); 4 - Strategic Information and Knowledge in Organizations (39); 5 - Information and Communication Technologies - ICT (100); 6 - Communication and Access to Information (96); 7 - Document and Information as Component (35); 8 - Areas of Knowledge (65). This profile is shown in Figure 4.

On the other hand, the profile data of the journal according to the facets of *TCI* controlled vocabulary are: 1 - BI - Information and library operations (73); 2 - CO - Knowledge and information (15); 3 - TI - Information Technology (64); 4- AR - Archival Science (15); 5 - FI - Information Sources (13); 6 - CD - Fields and disciplines (47); 7 - CI - Information Science (83); 8 - DC - Documentation (17); 9 - PP - People, professionals and formal groups (38); 10 - PM - Research and analytical methods (19); 11 - MU - Museology (1); 12 - UI - Information Units (48); 13 - OR - Organizations (5); 14 - MC - Physical and communication media (0).

Figure 4 - Informação & Informação Profile - Keyword Compatibilization with TBCI.



Source: by the authors, using SEER system layout

Thus, the journal has publications in the following categories: Information Science Epistemology, and Information Technology and Communication on its first profile; on its second profile, the journal has more publications in CI – Information Science and BI – Information and library operations.

Conclusions

The control of keyword vocabulary in electronic scientific articles, the compatibilization between the terms of controlled vocabularies and keywords, were carried out according to the parameters established in the methodological proposal, and within the validation expectations. To execute each step of VCI&I tools process, it was important to note the diversity of syntax in the keywords assigned to electronic scientific articles.

The index of terms provided by the tool becomes a means to acknowledge the journal in a holistic way. The journal's profile is another option to visualize the content it treats and its biases of scientific production. In addition, this visualization of Informação & Informação aim to assist authors, through the index of terms, in assigning keywords to their scientific productions.

With the development of VCI&I tools, it will be possible to apply the procedure to all journals in the area of Information Science, managed by SEER system. It is also hoped, for future studies, to make the tool applicable to other specialized areas managed by this publishing system.

References

The references was made following the ABNT rules.

AMERICAN NATIONAL STANDARDS INSTITUTE - ANSI; NATIONAL INFORMATION STANDARDS ORGANIZATION - NISO. *Z39:19-2005: guidelines for the construction, format, and management of monolingual controlled vocabularies*. Bethesda: NISO Press, 2005.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. *NBR 6028*. Informação e documentação: resumo: apresentação. Rio de Janeiro: 2003b.

FUJITA, M. S. L. Modelos de categorização para a construção de tesouros: metodologia de ensino. In: BOCCATO, Vera Regina Casari; GRACIOSO, Luciana de Souza (Org.). *Estudos de linguagem em ciência da informação*. Campinas: Alínea, 2011. Cap. 2.

GIL LEIVA, I.; ALONSO ARROYO, A. La relación entre las palabras clave aportadas por los autores de artículos de revista y su indexación en las bases de datos ISOC, IME e ICYT. *Revista Española de Documentación Científica*, Madrid, v. 28, n. 1, 2005. Available at: <<http://redc.revistas.csic.es/index.php/redc/article/view/165>>. Access at: 9 set. 2014.

GONÇALVES, A. L. Uso de resumos e palavras-chave em ciências sociais: uma avaliação. *Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação*, Florianópolis, v. 13, n. 26, p. 1-15, 2008. Available at: <<https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2008v13n26p78/6636>>. Access at: 9 set. 2014.

MIGUÉIS, A.; NEVES, B. Uma abordagem à linguagem de indexação dos artigos científicos depositados no repositório científico da Universidade de Coimbra. *Pontodeacesso*, Salvador, v. 7, n. 1, p. 116-131, abr. 2013.

MOREIRA, M. P. *Tesouro em Ciência da Informação*. Belo Horizonte: PUC Minas. Instituto de Informática; UFMG, [2006]. Disponível em: <http://icei.pucminas.br/ci/tci/index.php>. Acesso em: 9 mar. 2015.

ORENGO, V. M.; HUYCK, C. A Stemming Algorithm for the Portuguese Language. In: 8th International Symposium On String Processing And Information Retrieval (spire), Laguna de San Raphael, Chile, 2001, *Proceedings...* p.183-193, DOI: 10.1109/SPIRE.2001.989755

ORENGO, V. M.; BURIOL, L. S.; COELHO, A. R. A Study on the Use of Stemming for Monolingual Ad-Hoc Portuguese Information Retrieval. In: CLEF 2006, Alicante, Berlin, *Proceedings...* Heidelberg: Springer-Verlag. v. 4730 p. 91-98. DOI: 10.1007/978-3-540-74999-8_12.

PINHEIRO, L. V. R.; FERREZ, H. D. *Tesouro Brasileiro de Ciência da Informação*. Rio de Janeiro; Brasília: Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict), 2014. 384 p. Available at: <http://www.ibict.br/publicacoes-e-institucionais/tesouro-brasileiro-de-ciencia-da-informacao-1/tesouro-brasileiro-de-ciencia-da-informacao/resolveuid/1c60ede36f47aee60c48957ef6db7510>. Access at: 12 nov. 2014.

Documentary Languages in archival description



Maria de Fátima Santos de Lima
Librarian
fatimasl2006@yahoo.com.br

Francisco Aragão Pedroza da Cunha
Federal University of Bahia
pedrozaici@gmail.com

Introduction

This article is the result of the Conclusion Paper of the Undergraduate Program in Archival Science by Federal University of Bahia presented in 2014.2, which had, as its theme, the information representation in archives from Archival Description.

Archival Description is an essential research process for information treatment, representation, retrieval and consequently, as a tool for the information dissemination. Therefore, it is a key activity to develop another activity within the scope of Archival Science - information dissemination.

Thus, this paper aims to verify the use of taxonomies and ontologies to represent organic information contained in the documents kept by state public archival institutions in Brazil.

The nature of the research was applied with a qualitative and quantitative approach. According to the objectives, it was classified as exploratory and descriptive. The universe was the Brazilian state public archival institutions, and the selected sample were the institutions that have official websites and offer search tools for online consultation.

Archival description

To represent a collection, it is necessary to develop activities and processes that effectively allow information use and dissemination. In this sense, Archival Description activities can be considered as a "process by which the elaborator of the description apprehends, identifies, condenses and, without distortions, presents all possibilities of use and application of the documentation described by him" (BELLOTTO , 2006, p 178, our translation).²⁵⁸

Generally, Archival Description is related only to the elaboration of research instruments. As Oliveira (2010, 43) points out, "archival description has been presented in the literature in a reductive manner in relation to what it actually means and represents, [...] in the last decades, it is linked to compliance with standards which aim only at the elaboration of research instruments [...]." However, in addition to the elaboration of a research instrument, Archival Description establishes intellectual control, explains and promotes access to collections, and creates means to represent information according to the context of document creation.

In addition to considering the document creation context, Archival Description activity covers a series of processes and activities, such as: research, indexing, organization, elaboration of research tools, information representation, knowledge representation as a goal to disseminate information.

It is then time to define information representation and knowledge representation. Thus, information representation can be understood as the "set of attributes that represents a given informational object and that is obtained by processes of physical description and document content; and knowledge representation is a conceptual structure that represents world models and can be performed through different types of knowledge organization systems (KOS)" (BRASCHER; CAFÉ, 2008, p. 6). My choice in this research is the information representation from knowledge organization systems.

For Dodebei (2002), information representation is composed of two elements: objects (i.e. things we want to represent), which, in the context of this paper are the documents; and properties (i.e., characteristics of such things), the words or descriptors by which the documents will be represented.

Information representation through documentary languages supports the description process and the creation of products to help access and retrieve this information. "These languages are therefore constructed for indexing, storing and retrieving information and correspond to symbol systems intended to 'translate' the contents of documents" (CINTRA, et.al.2002, p.16). As such, they facilitate communication, represent information in a synthetic way, and provide faster search and retrieval. ²⁵⁹

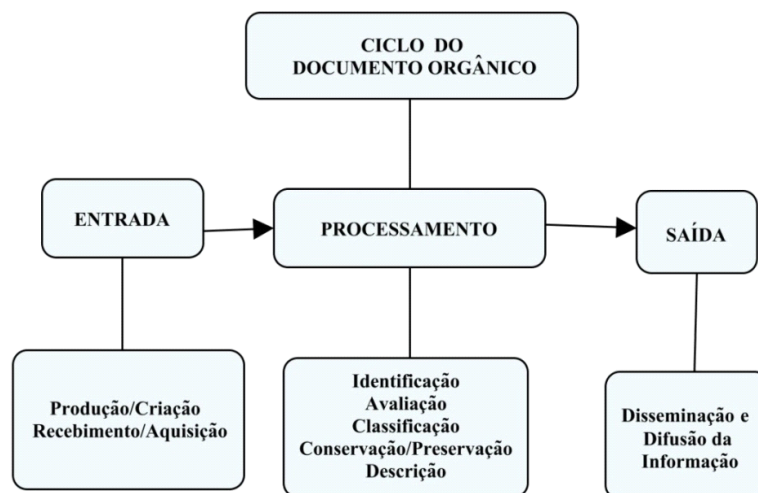
Thematic treatment of organic information

Thematic treatment of information can be understood as the process that involves physical and content description of informational objects, the product of this descriptive process is information representation, understood as a set of descriptive elements that represent the attributes of a specific informational object (BRASCHER, CAFÉ, 2008).

Thematic treatment of information is responsible for mediation between information producer and user, exercising the fundamental task of giving access to documentary content (NARUKAWA; SALES 2012, p.154). This treatment is part of a set of operations named documentary cycle.

According to Robredo (2005, p.8), documentary cycle "can be considered as a system which incorporates a series of elements to be treated and converted into a new product, easier to disseminate or to be assimilated by the user." This cycle involves three basic processes: input, processing, and output. For the same author, input consists of the selection and acquisition of document, the processing is the set of activities related to the technical treatment of the document, and output consists of information search and retrieval.

In the context of the organic or archival document, this cycle is linked to archival activities which, for Rousseau and Couture (1998), are named functions and are related to the set of principles and operations that concern the treatment and organization of archives. For these theorists, these functions are the creation, acquisition, evaluation, conservation, classification, description and diffusion. It is worth noting that Rodrigues (2011) adds another function: identification. Figure 1 illustrates the organic document

Figure 1: Organic Document Cycle.

Source: by the author, based on Robredo (2005)

In the organic document cycle, input can be defined as the production/creation stage in which organizations, people or producing families generate documents for a variety of purposes. On the other hand, the receipt or acquisition takes place when documentary collections are collected or transferred to archival institutions or to historical archives, thus categorizing, document custody organizations.

All steps in this process are important. However, only the Description is object of this study. In Archival Description activity, it is worth highlighting the importance of using documentary languages - ontologies and taxonomies - during the indexing process, as it contributes to the quality of the results of organic documents access and retrieval operations.

Indexing in the area of Information Science can be understood as the process by which the indexer identifies concepts that a particular document treats and expresses through descriptors or keywords. For Robredo (2005, p. 165), indexing consists of indicating the thematic content of an information unit, by assigning one or more terms (or codes) to the document, univocally.

There are two ways of indexing documents: indexing by words and indexing by concept. The first one is to use the author's own terms, words found in the title or in the

text of documents, that is, natural language. The second one, indexing by concept, is carried out from words and their associations. The concepts are identified and, thus, they are codified or expressed in the form of descriptors, configuring controlled language (ROBREDO, 2005).

Indexing by concepts requires a greater degree of cognition from the indexer and longer for the completion of the process. This type of indexing extends the possibilities of information retrieval, as it standardizes the terminology and guarantees greater precision and effectiveness in information search. ²⁶¹

In archives, the indexing process is part of Archival Description, which, in turn, observes the characteristics and contexts of the documentary sets, that is, "the unit of documentary analysis will be a fund; series and subseries, respecting the Principle of Provenance." (AGUIAR, 2008, p. 184).

Principle of Provenance and Original Order ("Principle according to which the archive must preserve the arrangement given by the collective entity, person or family that produced it" (ARQUIVO NACIONAL, 2005, p. 137)), direct the entire process of documentary treatment in archives.

However, despite all the efforts of Archival Description, document search and retrieval are not satisfactory. The lack of accuracy and high revocation of descriptors assigned to documents during the indexing process, and the increase in the supply of collections through portals and web sites requires a vocabulary standardization. For Souza and Alvarenga (2004, p. 133), the web was implemented in an almost anarchic and chaotic way. These authors state that, on the web:

[...] there is no comprehensive and satisfactory strategy for indexing documents contained therein, and information retrieval, made possible through "search engines", is based primarily on keywords contained in the original document text, which is very ineffective, (SOUZA; ALVARENGA, 2004, p 133).

Polysemy and ambiguity of terms used in document indexing are other factors that make it difficult to retrieve information on the web due to the lack of use of semantic standards. Thus, semantic web emerges:

[...]semantic web is an extension of the current web, but it presents a structure that allows the understanding and management of contents

stored on the web regardless of the way in which they appear, be it text, sound, image and graphics from the semantic valuation of these contents, and through agents that will be content-collecting programs from diverse sources capable of processing information and exchanging results with other programs (BERNERS-LEE, 2001, *apud* RAMALHO, 2007, p. 3).

By adding semantics to a content, semantic web will determine the meaning of a term in the context of a particular document. Therefore, it facilitates the construction of applications aimed at reducing difficulties related to information retrieval and creates a cooperation environment (PICKLER, 2007).

This cooperation environment may be possible with the adoption of standardized concepts and control of the terminologies used in several areas of knowledge; "In order to facilitate information retrieval, research is being conducted over the last few years around the terms taxonomy and ontology. Experts from diverse fields of knowledge propose dynamic forms of information organization [...]" (NOVO, 2010, page 132).

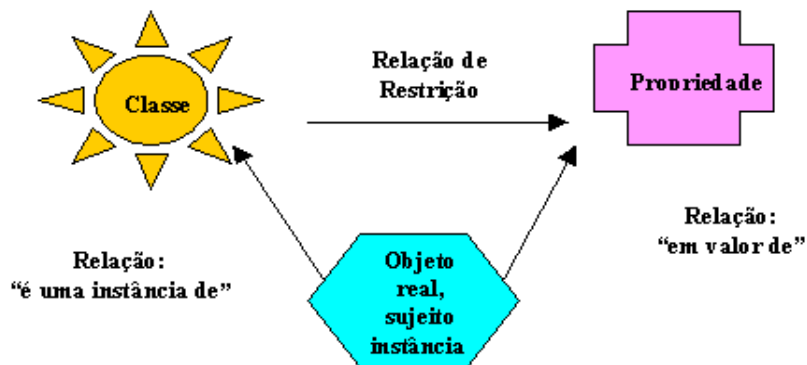
Ontologies are also used in Artificial Intelligence, Computer Science, Semantic Web and Information Science as knowledge representation and information. "In information representation, it can be understood as the sum of a series of relevant concepts that represent the knowledge shared by the members of a given domain" (MOREIRO GONZÁLEZ, 2011, p.76).

The use of ontologies facilitates information sharing, defines basic concepts and the relationship between them, defines the terms to describe and represent a particular area of knowledge and assigns meaning to document content.

Ontology is a set of standardized concepts where terms and definitions must be accepted by a community within a domain and are intended to allow multiple agents to share knowledge. (CAMPOS, 2007, p.2).

One of the main objectives of ontologies is the exchange of knowledge and its representation. For this, the concepts need to be contextualized and related. For Rodríguez Barquín et al. (2006), ontologies have some characteristics that are related to each other as shown in Figure 2:

Figure 2 - Relationships among classes, properties and instances



Source: Rodríguez Barquín *et. al.* (2006, p. 6).

It is inferred that classes are represented by concepts according to the analyzed context. Instances are the values assigned to a concept in a specific situation. And properties are the characteristics of concepts. In this sense, ontologies are composed of hierarchically structured concepts and have similar characteristics to taxonomies.

Like ontologies, taxonomies are also used to represent information. In the field of Information Science, taxonomy is always related to information organization and representation.

For Currás (2010, page 68), taxonomies are "[...] conceptual frameworks suitable for use in semantic web and as [the term] that creates order in the chaos of conceptual units, coming from extremely varied documents."

Taxonomy is a tool that allows classifying and facilitating access to information in a disorganized and disorderly environment such as the internet. For Moreira González (2011), in the field of documentary languages, it is a list of preferred terms with hierarchical structure (tree), which uses the parent-child relationship.

A hierarchical structure gathers terms of a certain area of knowledge that has relations with each other and each term is superior to the next term. The fund is related to the section, which, in turn, is related to the series and they all come from the same fund.

Both ontologies and taxonomies define concepts in knowledge domains to assist information representation contained in documents, in order to promote information retrieval and access. However there are some differences between them, as listed in Table 2:

Table 2: Basic differences between ontologies and taxonomies

CATEGORIES	ONTOLOGIES	TAXONOMIES
Relations among concepts	Applies basically to hierarchical relations.	The relations take place in a conceptual network.
Objectives	Seek the development of categories to represent knowledge and help information retrieval.	Seek linguistic consensus in specific areas to represent knowledge and help information retrieval.
Application	Are more indicated to organize and represent information in specific environments: portals e intranets.	Are more indicated to areas of knowledge that wish to work themes specifically and deeply.
Type of relationship among terms	“Generalization” or “type of”	“Part of”, “cause effect”, “location”, etc.

Source: by the author, based on Vital and Café (2011)

The network relationships, characteristic of ontologies give the conceptual structure more quality; from the semantic point of view, they are more improved and require a more complex intellectual level with the support of experts from their respective areas of knowledge so that the concepts are shared and understood by all. And in general, ontologies are considered to be more complex than taxonomies, since they represent a certain domain of knowledge (VITAL; CAFÉ, 2011).

In this perspective, it is worth mentioning that the use of ontologies and taxonomies in indexing systems is fundamental for the search, representation and retrieval of organic information. However, these systems of knowledge organization are still little used in the area of archival science.

Conclusions

There are many aspects to be considered for efficient search and retrieval of organic document information. In this context, Archival Description is consolidated as an essential research process for the treatment, representation, retrieval and consequently, as a tool for disseminating information, in addition to enabling the information exchange, it guarantees the Archive its role of knowledge disseminator.

Thus, Archival Description should not only be seen as a set of procedures for elaborating research instruments, but rather seen to enhance the recurrent use of

information in archives, and the use of information organization and representation tools during the indexing process, such as ontologies and taxonomies.

The results indicated that there is a latent need for more studies regarding ontologies and taxonomies as tools to organize and represent knowledge and information in KOS, and especially the applications in indexing systems in the area of Archival Science. ²⁶⁵

We also perceive the possibility of applying ontologies and taxonomies, as fundamental tools to organize and represent information in KOS in the field of Archival Science, and that provide the end user information retrieval quickly and safely.

References

The references was made following the ABNT rules.

AGUIAR, F. L. *O controle de vocabulário controlado como dispositivo metodológico para organização, tratamento e recuperação da informação arquivística*. 267 f. 2008. Dissertação (Mestrado em Ciência da Informação) – Pontifícia Universidade Católica de Campinas - PUC, Campinas, 2008. Available at: <http://www.bibliotecadigital.puc-campinas.edu.br/tde_busca/arquivo.php?codArquivo=437>. Access at: 10 set. 2014.

ARQUIVO NACIONAL. *Dicionário brasileiro de terminologia arquivística*. Rio de Janeiro: Arquivo Nacional, 2005. Disponível em: <http://www.arquivonacional.gov.br/Media/Dicion%20Term%20Arquiv.pdf>. Acesso em: 10 maio. 2014.

BELLOTTO, Heloísa Liberalli. O sentido de descrição documental. In: _____. *Arquivos permanentes: tratamento documental*. 4. ed. Rio de Janeiro: FGV, 2006, p. 173-178.

BRASCHER, M.; CAFÉ, L. Organização da informação ou organização do conhecimento? In: Encontro Nacional de Pesquisa em Ciência da Informação, 9., 2008. *Anais...* São Paulo: ANCIB, 2008. Disponível em: <[http://skat.ihmc.us/rid=1KR7TM7S9-S3HDKP-5STP/BRASCHER%20CAF%3%89\(2008\)-1835.pdf](http://skat.ihmc.us/rid=1KR7TM7S9-S3HDKP-5STP/BRASCHER%20CAF%3%89(2008)-1835.pdf)>. Acesso em: 28 out. 2014.

CAMPOS, Maria Luiza de Almeida. *A problemática da compatibilização terminológica e a integração de ontologias: o papel das definições conceituais*. Grupo Temático Ancib: GT 2: Organização do Conhecimento e Representação da Informação. Disponível em: <<http://www.uff.br/ppgci/editais/marialuizaleituras.pdf>>. Acesso em: 29 set. 2014.

CINTRA, Ana Maria. et. al. *Para entender as linguagens documentárias*. 2. ed. São Paulo: Polis, 2002.

CURRÁS, Emília. *Ontologias, taxonomia e tesaurus: em teoria dos sistemas e sistemática*. Brasília: Thesaurus, 2010.

DODEBEI, Vera Lúcia Doyle. *Tesouro: linguagem de representação da memória documentária*. Rio de Janeiro: Interciência: Niterói: Intertexto, 2002.

MOREIRO GONZÁLEZ, José Antonio. *Linguagens documentárias e vocabulários semânticos para a web: elementos conceituais*. Salvador: EDUFBA, 2011.

NARUKAWA, Cristina Miyuki; SALES, Rodrigo. A relevância e influência de linguagens documentárias na política de tratamento da informação. In: Gil Leiva, Isidoro; Fujita, Mariângela Spotti Lopes (Org.). *Política de indexação*. São Paulo: Cultura Acadêmica; Marília: Oficina Universitária, 2012. Available at: http://www.marilia.unesp.br/Home/Publicacoes/politica-de-indexacao_ebook.pdf. Access at: 10 set. 2014. | 266

NOVO, Hildenise Ferreira. A taxonomia enquanto estrutura classificatória: uma aplicação em domínio de conhecimento interdisciplinar. *Ponto de Acesso*, Salvador, v. 4, n. 2, p. 131-156, set. 2010. Available at: <http://www.portalseer.ufba.br/index.php/revistaici/article/viewArticle/4103>. Access at: 10 out. 2014.

OLIVEIRA, Lúcia Maria Velloso de. *Modelagem e status científico da descrição arquivística no campo dos arquivos pessoais*. 2010, 188f. Tese (Doutorado em História Social). Universidade de São Paulo – Faculdade de Filosofia, Letras e História Social – Departamento de História. São Paulo, 2010. Available at: <http://www.teses.usp.br/teses/disponiveis/8/8138/tde-14062011-134720/pt-br.php>. Access at: 15 set. 2014.

PICKLER, Maria Elisa Valentim. Web semântica: ontologias como ferramentas de representação do conhecimento. *Perspectivas em Ciência da Informação*. São Paulo, v. 12, n. 1, p. 65-83, jan./abr. 2007. Available at: www.brapci.ufpr.br/download.php?dd0=15365.> Access at: 10 out. 2014.

RAMALHO, Rogério Aparecido Sá. et. al. Web semântica: uma investigação sob o olhar da Ciência da Informação. *Data Grama Zero - Revista de Ciência da Informação*, v.8, n.6 dez/07. Disponível em: http://www.dgz.org.br/dez07/Art_04.htm. Acesso em: 10 out. 2014.

ROBREDO, Jaime. *Documentação de hoje e de amanhã: uma abordagem revisitada e contemporânea da Ciência da Informação e de suas aplicações biblioteconômicas, documentárias, arquivísticas e museológicas* 4ª ed. Brasília: Edição do autor. 2005.

RODRIGUES, Ana Célia. Identificação: uma nova função arquivística?. *Revista EDICIC*, v.1, n.4, p.109-129, Out./Dez. 2011. Available at: <http://www.edicic.org/revista/>.> Access at: 10 set. 2014

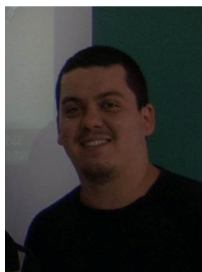
RODRÍGUEZ BARQUÍN, Beatriz Ainhize; MOREIRO GONZÁLEZ, José Antonio; PINTO, Adilson Luiz. Construção de uma ontologia para sistemas de informação empresarial para a área de Telecomunicações. *Data Grama Zero: revista de Ciência da Informação*, v.7 n.2, abr. 2006. Available at: http://www.dgz.org.br/abr06/Art_04.htm> Access at: 30 set. 2014.

ROUSSEAU, Jean-Yves; COUTURE, Carol. O lugar da arquivística na gestão da informação. In: _____. *Os fundamentos da disciplina Arquivística*. Lisboa: Dom Quixote, 1998.

SOUZA, Renato Rocha; ALVARENGA, Lídia. A Web Semântica e suas contribuições para a Ciência da Informação. *Ci. Inf.*, Brasília, v. 33, n. 1, p. 132-141, jan./abril 2004. Available at: <<http://www.scielo.br/pdf/ci/v33n1/v33n1a16.pdf>> . Access at: 01 out. 2014.

VITAL, Luciane Paula; CAFÉ, Ligia Maria Arruda. Ontologias e taxonomias: diferenças. *Perspect. ciênc. inf.*, Belo Horizonte, v.16 n. 2, abr./jun. 2011. Available at: <http://www.scielo.br/scielo.php?pid=S1413-99362011000200008&script=sci_arttext > Access at: 14 set. 2014. 267

Research in knowledge organization systems in DCMI conferences



Felipe Augusto Arakaki
São Paulo State University
fe.arakaki@gmail.com



Plácida Leopoldina Ventura Amorim da Costa Santos
São Paulo State University
placida@marilia.unesp.br



Rachel Cristina Vesú Alves
São Paulo State University
rachel@marilia.unesp

Introduction

With the changes in media caused by technologies, Knowledge Organization has always been willing to discuss new possibilities and improvements for organization and representation of resources on the Web environment.

In this sense, much of its researchers contributed to the development of Dublin Core, a metadata standard created in 1995 to locate resources on the Web environment. Since the creation of Dublin Core, researchers from various fields have gathered at conferences promoted by the Dublin Core Metadata Initiative (DMCI), which administers the Dublin Core, to discuss metadata issues on the Web environment.

Thus, with the proposal of internationalizing the Dublin Core standard, from 2001, the current structure began to accept paper presentations and many were the research contributions concerning Knowledge Organization Systems (KOS).

Knowledge Organization System (KOS), which encompasses all types of schemas that organize and represent knowledge, e.g. classifications, taxonomies, thesauri and ontologies. KOS are semantically structured conceptual systems that contemplate terms, definitions, relationships and properties of concepts. (CARLANZ 269 MEDEIROS, 2011, p. 54).

In that period, from 2001 to 2014, 14 conferences were held in total, five (5) in Europe and five (5) Asia, three (3) in North America and one in Central America.

In this paper, we question, what is the scenario of KOS-related research presented in DCMI events? In this way, the objective is to present the state of the art of the studies presented in conferences organized by DCMI that somehow approach the theme of KOS.

Methodology

It is a theoretical research, qualitative and bibliographical, in order to identify the state of the art of the research presented in the DCMI events from 2001 to 2014 in English language.

Only papers that deal with KOS thematic studies such as thesauri, ontologies, controlled vocabularies, and others were selected. The steps were: I. Identification of complete papers presented in DCMI (380 papers); II. Selection of papers from the title, keywords and abstract (35 papers); III. Exploratory reading of texts; IV. Extraction of thematic; V. Classification of research categories; VI. Outline of the study and analysis of the results.

In the years 2001 to 2004, 2010 and 2014, no papers were identified that addressed the subject in question.

Research in Knowledge Organization Systems at DCMI conferences

Among the contributions and influences of Knowledge Organization in DCMI, we highlight, for example, the conference held in Madrid in 2005, which addressed the main theme "Vocabularies in Practice". During the conference in Spain, several perspectives on the implementation of controlled vocabularies were exposed, as well as suggestions and efforts to standardize them.

Miles et al. (2005) discussed a terminological definition for controlled vocabulary. Lee and Sugimoto (2005) tried to standardize controlled vocabularies in three domains. Greenberg et al. (2005) discussed some principles on controlled vocabulary. Pérez Agüera and Sánchez Jiménez (2005) explored the initiative of inserting a thesaurus in the Resource Description Framework (RDF) model. Francesconi and Peruginelli (2005)¹²⁷⁰ defined a standardization of metadata and controlled vocabulary for Italian legal articles. Kurth, Nehler, and Silterra (2005) explored the particular case of a library by administering a controlled vocabulary.

The following year, in 2006, Colima in Mexico addressed as theme "Metadata for Knowledge and Learning". It presented questions focused mainly on the standardization and control of vocabularies. Ángel Marzal, Ruiz Colmenero and Cuevas Cerveró (2006) proposed a controlled vocabulary for the description of educational objects. The issues of controlled environmental vocabulary sharing are addressed by Menger and Rütther (2006) who reported on a service of semantic network, developed by the Umweltbundesamt (UBA), German Federal Environment Agency and is integrated by several environmental information systems and Geographical area. Currier et al. (2006) discussed the need to develop and use vocabularies that describe educational objects in the United Kingdom context. Bartolo et al. (2006) addressed the development of the National Science Digital Library (NSDL) controlled vocabulary to facilitate catalog feeding from the Dublin Core metadata capture during computer code execution. Miles (2006) proposed a review of Simple Knowledge Organization System (SKOS) project prior to a standardization initiative and presented three issues regarding the development of SKOS for World Wide Web Consortium (W3C) recommendation. Harper (2006) addressed the EXtensible Stylesheet Language Transformations (XSLT) style sheets to convert *Library of Congress Headings for Subject LCSH Authorities* based on Machine Readable Cataloging eXtensible Markup Language (MARCXML) or Metadata Authority Description Schema (MADS) into eXtensible Markup Language (XML) for RDF documents according to the SKOS project. In this same conception, Liang et al. (2006) addressed the conversion of AGROVOC data to a new web-based Ontology Language (OWL) system. Pattuelli and Norberg (2006) advocated an Ontologydriven that would represent social, familial, spatial

and temporal relationships among learning objects to facilitate the construction of historical meanings.

With the theme "Application Profiles and their Application in Practice", Singapore hosted the conference in 2007. Kakali et al. (2007) proposed an ontology for the application profile in museology. Miyazawa (2007) questioned the current language model in information sharing and the need for standardization of linguistic representation. Campbell (2007) explored the process of identifying things and how to couple from similar characteristics.

The following year, with papers addressing issues such as application profile and folksonomy, they were presented in Berlin, with the central theme "Metadata for Semantic and Social Applications" in 2008. Catarino, in partnership with Baptista (2008), proposed an application profile based on folksonomy. With a similar theme, Chen, Liu and Qin (2008) addressed a methodology for generating metadata based on tag extraction. Kim et al. (2008) mapped the state of the art of the ontology of tags. Furthermore, Lauser et al. (2008) discussed the creation and elaboration of thesauri by humans and not humans (automatically). Summers et al. (2008) discussed a tool that transforms an LCSH subject header into MARCXML for SKOS vocabulary in RDF.

The city of Seoul, South Korea, hosted the theme "Semantic Interoperability of Linked Data" in 2009. Mason (2009) addressed knowledge management and Dublin Core and the structural relationships of conceptual models of knowledge organization systems that influence the effectiveness and flow of knowledge. Panzer and Zeng (2009) proposed a classification modeling recommendation in the SKOS system. Jeong (2009) advocated using tags to aid in the representation of videos on YouTube.

Later, the event took place in the city of Hague, in the Netherlands, and addressed the theme "Metadata Harmonization: Bridging Languages of Description" in the year 2011. In the thematic field, Morshed et al. (2011) worked with thesaurus on Linked Data. Freire, Borbinha and Caldo (2011) sought the alignment of SKOS with a geographic ontology. Montiel-Ponsoda et al. (2011) addressed a standardized approach to nomenclature and labeling of ontologies for various languages on the Web. Zavalina (2011) compared free-texts and collection level of subjects at large scale in digital libraries.

In the year 2012, from Kuching, Sarawak, Malaysia with the theme "Metadata for Meeting Global Challenges". Among the papers presented, Ambiah and Lukose (2012) proposed a tool to automatically generate pages with semantic information from keywords in the document. Zumer, Zeng and Hlava (2012) presented a domain model to describe and access SKOS resources. Ontologies were approached by Bosch et al. (2012) who²⁷² addressed the development of a data documentation initiative, thus defining an international standard for data documentation and management in the area of Social, Behavioral and Economic Sciences.

The city of Lisbon in Portugal hosted the theme "Linking to the Future" in 2013. The paper by Tian et al. (2013) discussed an application profile and ontology based on Linked Data for Chinese traditional music. Hillmann, Dunsire, and Phipps (2013) verified the terminology changes of various metadata patterns. In this same line, Honma, Nagamori and Sugimoto (2013) addressed the location and combination of vocabularies to be projected in application profiles.

Analysis of research by categories

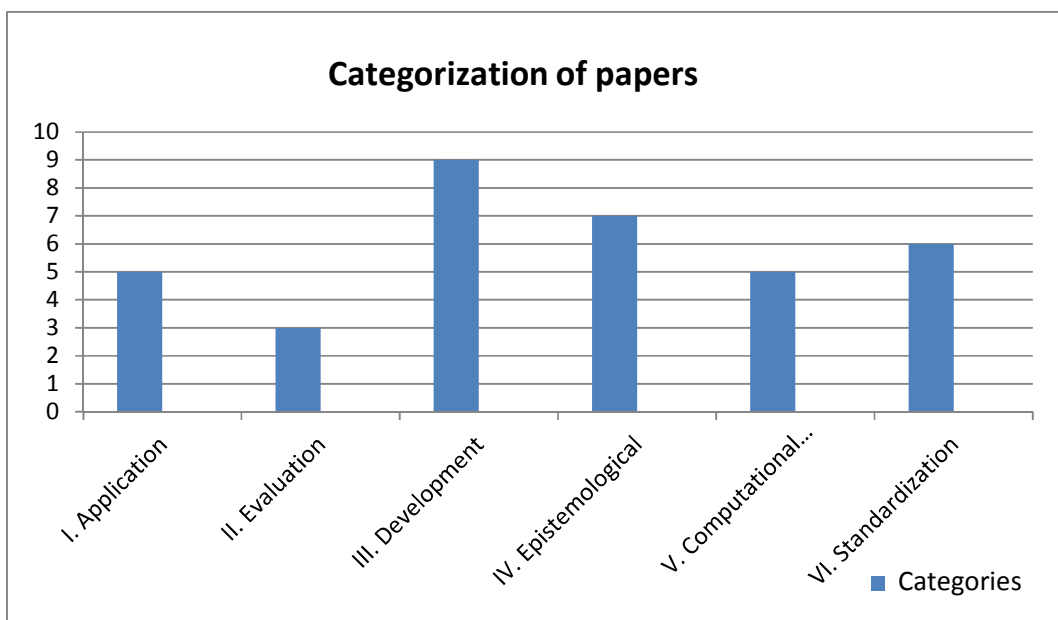
From the nature of the papers, we classified them into five categories:

- I. **Application** - Proposals or applications of controlled vocabularies, thesauri, ontologies, etc. in digital environments and institutions.
- II. **Evaluation** - Studies that seek to evaluate the effectiveness and efficiency of controlled vocabularies in systems.
- III. **Development** - Proposals for the development of controlled vocabularies, ontologies, thesauri, etc. in digital environments.
- IV. **Epistemological** - Research that approached and reflected questions generating concepts, foundations, techniques, studies, explanations and criticisms about Knowledge Organization Systems.
- V. **Computational Structures** - Papers that presented implementation of computational languages such as RDF, OWL, etc. in controlled vocabularies, thesauri, etc.

VI. **Standardization** - Initiatives to standardize controlled vocabularies, terminologies, ontologies, thesauri, etc.

Figure 1 represents the mapping of the papers classified in the presented categories.

Figure 1. Number of papers in relation to the categories

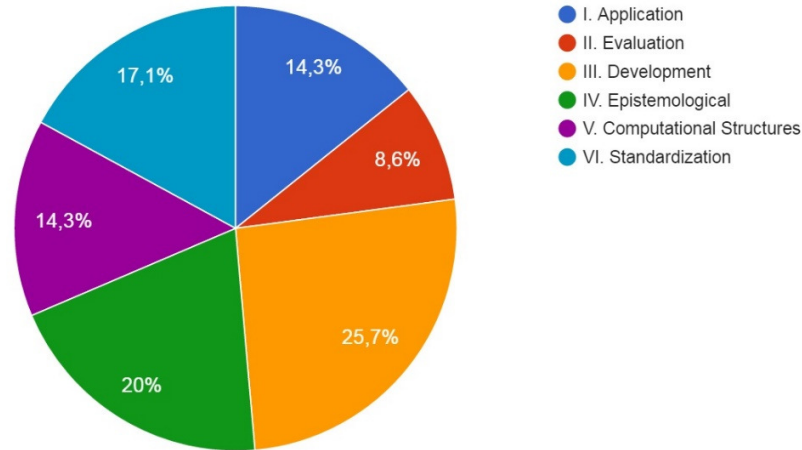


Source: by the authors

According to Figure 1, in category I. Application five (5) papers were classified, representing 14.3% of the total research; category II. Evaluation, presented three (3) papers, or 8.6%; category III. Development stands out with nine (9) papers, totaling 25.7%; category IV. Epistemological analysis presents 20% of the studies, that is to say, seven (7) papers, and category V. Computational structures presented five (5) papers, or 14.3%, and finally, category VI. Standardization, with six (6) papers presented, representing 17.1% of the total papers. Figure 2 shows the distribution of the percentages of the papers.

Figure 2. Distribution by percentage of papers

Percentage of papers



Source: by the authors

As observed, the research presented in DCMI conferences that approached Knowledge Organization Systems have the greatest nature in proposals for their development (category III.); secondly, the epistemological studies (category IV.), thirdly the issues involving standardization (Category VI), followed by applications tied with structure (categories I. and V. respectively) and lastly research regarding evaluation (category II).

Conclusions

From the analysis of the research presented in DCMI conferences, it can be observed that the panorama of studies in Knowledge Organization Systems can be divided into 6 categories: application, evaluation, development, epistemological studies, computational structures and standardization. The state of the art revealed that most studies are focused on the construction and development of thesauri, ontologies and controlled vocabularies, category 3; followed by epistemological studies, which together

accounted for over 45% of the research presented. Few studies were related to the evaluation of Knowledge Organization Systems with approximately 8.6% of the papers.

It is suggested, for further studies, the extension of the research cut in the abstracts presented in DCMI conferences, as well as the application and comparison of the study in other conferences of the area and journals. This analysis could broaden the panorama and reveal the trends of studies in Knowledge Organization Systems.

References

The references was made following the ABNT rules.

AMBIAH, N.; LUKOSE, D. Enriching webpages with semantic information. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 12., 2012, Sarawak. *Proceedings...* Malaysia: DCMI, 2012. Available in: <<http://dcpapers.dublincore.org/pubs/article/view/3681>>. Access in: 26 abr. 2015.

BARTOLO, L. et al. Development of a wiki-based, expert community-driven nanosystem vocabulary. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 6., 2006, Colima. *Proceedings...* Mexico: DCMI, 2006. Available in: <<http://dcpapers.dublincore.org/pubs/article/view/852>>. Access in: 16 abr. 2015.

BOSCH, T. et al. Leveraging the DDI model for linked statistical data in the social, behavioral, and economic sciences. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 12., 2012, Sarawak. *Proceedings...* Malaysia: DCMI, 2012. Available in: <<http://dcpapers.dublincore.org/pubs/article/view/3654>>. Access in: 23 abr. 2015.

CAMPBELL, D. Identifying the identifiers. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 7., 2007, Singapore. *Proceedings...* Singapore: DCMI, 2007. Available in: <<http://dcpapers.dublincore.org/pubs/article/view/868>>. Access in: 20 abr. 2015.

CARLAN, E.; MEDEIROS, M. B. B. Sistemas de Organização do Conhecimento na visão da Ciência da Informação. *RICI: R. Ibero-amer. Ci. Inf.*, Brasília, v. 4, n. 2, p. 53-73, ago./dez.2011. Available in: <<http://periodicos.unb.br/index.php/RICI/article/view/6209>>. Access in: 06 maio 2015.

CATARINO, M.; BAPTISTA, A. Relating folksonomies with Dublin Core. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 8., 2008, Berlin. *Proceedings...* Germany: DCMI, 2008. Available in: <<http://dcpapers.dublincore.org/pubs/article/view/915>>. Access in: 21 abr. 2015.

CHEN, M.; LIU, X.; QIN, J. Semantic relation extraction from socially-generated tags: a methodology for metadata generation. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 8., 2008, Berlin. *Proceedings...* Germany: DCMI, 2008. Available in: <<http://dcpapers.dublincore.org/pubs/article/view/924>>. Access in: 21 abr. 2015.

CURRIER, S. et al. Vocabularies for describing pedagogical approach in e-Learning: a scoping study. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 6., 2006, Colima. *Proceedings...* Mexico: DCMI, 2006. Available in: <<http://dcpapers.dublincore.org/pubs/article/view/847>>. Access in: 20 abr. 2015.

FRANCESCONI, E.; PERUGINELL, G. Retrieval of Italian legal literature: a case of semantic search using legal vocabulary. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 5., 2005, Madrid. *Proceedings...* Spain: DCMI, 2005. Available in: <<http://dcpapers.dublincore.org/pubs/article/view/807>>. Access in: 15 abr. 2015.

FREIRE, N.; BORBINHA, J.; CALADO, P. A language independent approach for aligning subject heading systems with geographic ontologies. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 11., 2011, Hague. *Proceedings...* Netherlands: DCMI, 2011. Available in: <<http://dcpapers.dublincore.org/pubs/article/view/3624>>. Access in: 26 abr. 2015.

GREENBERG, J. et al. Growing vocabularies for plant identification and scientific learning. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 5., 2005, Madrid. *Proceedings...* Spain: DCMI, 2005. Available in: <<http://dcpapers.dublincore.org/pubs/article/view/808>>. Access in: 20 abr. 2015.

HARPER, C. Encoding Library of Congress Subject Headings in SKOS: Authority control for the Semantic Web. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 6., 2006, Colima. *Proceedings...* Mexico: DCMI, 2006. Available in: <<http://dcpapers.dublincore.org/pubs/article/view/842>>. Access in: 21 abr. 2015.

HILLMANN, D.; DUNSIRE, G.; PHIPPS, J. Maps and gaps: strategies for vocabulary design and development. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 13., 2013, Lisboa. *Proceedings...* Portugal: DCMI, 2013. Available in: <<http://dcpapers.dublincore.org/pubs/article/view/3673>>. Access in: 25 abr. 2015.

HONMA, T.; NAGAMORI, M.; SUGIMOTO, S. Find and combine vocabularies to design metadata application profiles using schema registries and LOD resources. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 13., 2013, Lisboa. *Proceedings...* Portugal: DCMI, 2013. Available in: <<http://dcpapers.dublincore.org/pubs/article/view/3675>>. Access in: 25 abr. 2015.

JEONG, W. Is tagging effective?: overlapping ratios with other metadata fields. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 9., 2009, Seoul. *Proceedings...* South Korea: DCMI, 2009. Available in: <<http://dcpapers.dublincore.org/pubs/article/view/975>>. Access in: 19 abr. 2015.

KAKALI, C. et al. Integrating Dublin Core metadata for cultural heritage collections using ontologies. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 7., 2007, Singapore. *Proceedings...* Singapore: DCMI, 2007. Available in: <<http://dcpapers.dublincore.org/pubs/article/view/871>>. Access in: 18 abr. 2015.

KIM, H. et al. The state of the art in tag ontologies: a semantic model for tagging and folksonomies. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 8., 2008, Berlin. *Proceedings...* Germany: DCMI, 2008. Available in: <<http://dcpapers.dublincore.org/pubs/article/view/925>>. Access in: 18 abr. 2015.

KURTH, M.; NEHLER, G.; SILTERRA, R. Using controlled vocabularies to manage resource relationships: the KMODDL experience. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 5., 2005, Madrid. *Proceedings...* Spain: DCMI, 2005. Available in: <<http://dcpapers.dublincore.org/pubs/article/view/811>>. Access in: 15 abr. 2015.

LAUSER, B. et al. Comparing human and automatic thesaurus mapping approaches in the agricultural domain. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 8., 2008, Berlin. *Proceedings...* Germany: DCMI, 2008. Available in: <<http://dcpapers.dublincore.org/pubs/article/view/918>>. Access in: 20 abr. 2015. | 277

LEE, W.; SUGIMOTO, S. Toward core subject vocabularies for community-oriented subject gateways. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 5., 2005, Madrid. *Proceedings...* Spain: DCMI, 2005. Available in: <<http://dcpapers.dublincore.org/pubs/article/view/769>>. Access in: 21 abr. 2015.

LIANG, A. et al. From AGROVOC to the Agricultural Ontology Service / Concept Server: An OWL model for creating ontologies in the agricultural domain. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 6., 2006, Colima. *Proceedings...* Mexico: DCMI, 2006. Available in: <<http://dcpapers.dublincore.org/pubs/article/view/841>>. Access in: 19 abr. 2015.

MARZAL, M.; RUIZ, M.; CERVERÓ, A. Development of a controlled vocabulary for learning objects' functional description in an educational repository. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 6., 2006, Colima. *Proceedings...* Mexico: DCMI, 2006. Available in: <<http://dcpapers.dublincore.org/pubs/article/view/857>>. Access in: 19 abr. 2015.

MASON, J. Knowledge management and Dublin Core. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 9., 2009, Seoul. *Proceedings...* South Korea: DCMI, 2009. Available in: <<http://dcpapers.dublincore.org/pubs/article/view/953>>. Access in: 21 abr. 2015.

MENGER, M.; RÜTHER, M. The Semantic Network Service: supporting heterogeneous environmental information systems. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 6., 2006, Colima. *Proceedings...* Mexico: DCMI, 2006. Available in: <<http://dcpapers.dublincore.org/pubs/article/view/860>>. Access in: 21 abr. 2015.

MILES, A. et al. SKOS Core: simple knowledge organisation for the web. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 5., 2005, Madrid. *Proceedings...* Spain: DCMI, 2005. Available in: <<http://dcpapers.dublincore.org/pubs/article/view/768>>. Access in: 21 abr. 2015.

MILES, A. SKOS: Requirements for standardization. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 6., 2006, Colima. *Proceedings...* Mexico: DCMI, 2006. Available in: <<http://dcpapers.dublincore.org/pubs/article/view/839>>. Access in: 11 abr. 2015.

MIYAZAWA, A. Parallel writing in east asian languages and its representation in metadata in light of the DCMI Abstract Model. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 7., 2007, Singapore. *Proceedings...* Singapore: DCMI,

2007. Available in: <<http://dcpapers.dublincore.org/pubs/article/view/863>>. Access in: 17 abr. 2015.

MONTIEL-PONSODA, E. et al. Style guidelines for naming and labeling ontologies in the multilingual Web. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 11., 2011, Hague. *Proceedings...* Netherlands: DCMI, 2011. Available in: <<http://dcpapers.dublincore.org/pubs/article/view/3626>>. Access in: 20 abr. 2015.

| 278

MORSHED, A. et al. Thesaurus alignment for Linked Data publishing. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 11., 2011, Hague. *Proceedings...* Netherlands: DCMI, 2011. Available in: <<http://dcpapers.dublincore.org/pubs/article/view/3623>>. Access in: 19 abr. 2015.

PANZER, M., ZENG, M. Modeling Classification Systems in SKOS: Some Challenges and Best-Practice Recommendations. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 9., 2009, Seul. *Proceedings...* South Korea: DCMI, 2009. Available in: <<http://dcpapers.dublincore.org/pubs/article/view/974>>. Access in: 13 abr. 2015.

PATTUELLI, M.; NORBERG, L. Identity matters: constructing social identities through ontology-based metadata. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 6., 2006, Colima. *Proceedings...* Mexico: DCMI, 2006. Available in: <<http://dcpapers.dublincore.org/pubs/article/view/853>>. Access in: 13 abr. 2015.

PÉREZ AGÜERA, J.; JIMÉNEZ, R. Thesauri managing and software agents: a proposed architecture. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 5., 2005, Madrid. *Proceedings...* Spain: DCMI, 2005. Available in: <<http://dcpapers.dublincore.org/pubs/article/view/810>>. Access in: 16 mar. 2015.

SUMMERS, E. et al. LCSH, SKOS and Linked Data. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 8., 2008, Berlin. *Proceedings...* Germany: DCMI, 2008. Available in: <<http://dcpapers.dublincore.org/pubs/article/view/916>>. Access in: 11 abr. 2015.

TIAN, M. et al. Towards the representation of Chinese traditional music: a state of the art review of music metadata standards. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 13., 2013, Lisboa. *Proceedings...* Portugal: DCMI, 2013. Available in: <<http://dcpapers.dublincore.org/pubs/article/view/3672>>. Access in: 13 abr. 2015.

ZAVALINA, O. Free-text collection-level subject metadata in large-scale digital libraries: a comparative content analysis. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 11., 2011, Hague. *Proceedings...* Netherlands: DCMI, 2011. Available in: <<http://dcpapers.dublincore.org/pubs/article/view/3630>>. Access in: 13 abr. 2015.

ŽUMER, M.; ZENG, M.; HLAVA, M. A domain model for describing and accessing KOS resources: report of processes in developing a KOS description metadata application profile. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 12., 2012, Sarawak. *Proceedings...* Malaysia: DCMI, 2012. Available in: <<http://dcpapers.dublincore.org/pubs/article/view/3656>>. Access in: 13 abr. 2015.

Conceptual models and descriptive representation of information



Prof. Dra. Elisabete Gonçalves de Souza
Federal Fluminense University
elisabetes.souza@gmail.com



Wellington Freire Cunha Costa
Federal Fluminense University
freirew@yahoo.com.br

Introduction

Metadata are essential for documentary description as they transform information into codes, assisting in their representation, retrieval, and access. Digital repositories make use of metadata to describe and allow online information search and retrieve, permitting database interoperability.

According to Marcondes (2005, p.96):

One of the major goals of using metadata in the Web context is, not only to describe electronic documents and information in general, but also to allow the evaluation of relevance by human users, as well as to allow special computers and programs, robots and software agents so they understand the metadata associated with documents and can, therefore, retrieve them, assess their relevance, and manipulate them more efficiently.

To ensure the integrity of system descriptions and interoperability, the use of rules and standards is paramount. For users, this action allows them to locate and choose among the various expressions and manifestations of a resource, the one that best meets their information needs. For managers, it means sharing records, using and reusing the information already processed by contextualizing it according to the policy of their information system.

Dublin Core (DC) is the most used standard for metadata structuring to create repositories and has the following characteristics: a) flexibility - the elements that compose

it are optional, can be displayed randomly and if need, be can be repeated or modified by qualifiers; b) simplicity - it is easy to handle, being self-explanatory, allowing the author of the document to make the description without the need of intense training; c) extensibility - because it is a simple and flexible model, it allows several communities in different areas to use the DC standard by exchanging information and obtaining access to them; d) interoperability - a common description model used by several areas gives users greater research skills; e) international scope - although it has been developed in English, there are already versions in other languages; proper infrastructure and international scope (DCMI, 2015).

According to Marcondes (2005) the DC standard was thought to be simple enough and self-explanatory so that the author of the document can describe it when publishing it electronically, which would be impossible in another standard, such as MARC, for example, as it requires the support of trained information professionals given its complexity. Going in the opposite direction,

Dublin Core can be defined as the set of metadata elements planned to facilitate the description of electronic resources. The expectation is that authors and Websites, which do not have cataloging skills, will be able to use Dublin Core to describe electronic resources, making their productions more visible to search engines and retrieval systems (SOUZA, VENDRUSCULO, MELO, 2000, p.93).

DC is a standard that complements the others, for its simplicity in the description of records, functioning as a "middle-term" between two types of description: the automatic ones performed by the Indexing and Summary Services and the catalogs produced by the library. Because it is based on the Z39.50 protocol (now restructured as SRU - Search / Retrieve URL, and SRW - Search / Retrieve Web Service), it allows the exchange of information between different database systems, simultaneous search and import of records, and can be deployed on any platform.

Interoperability of descriptions made in DC is guaranteed by the use of Extensible Markup Language (XML) and RDF (Resource Description Framework) semantics. According to Marcondes (2005, p. 108) "while XML is a generic language for describing electronic document structuring, RDF is specific for creating metadata for the purpose of locating and identifying resources. RDF coding allows to describe and search resources in a more flexible way, through the representation of relationships between them, in the

form of triples composed of object-attribute-value, where a value can be a literal or another resource" (CAMPOS; CAMPOS, 2005, p. 66). That is to say: through semantic interoperability, the DC format can be used by several areas and by being flexible, it allows extensibility and modularity of metadata, making it possible to exchange information between scientific communities easily and quickly.

Entity-relationship model

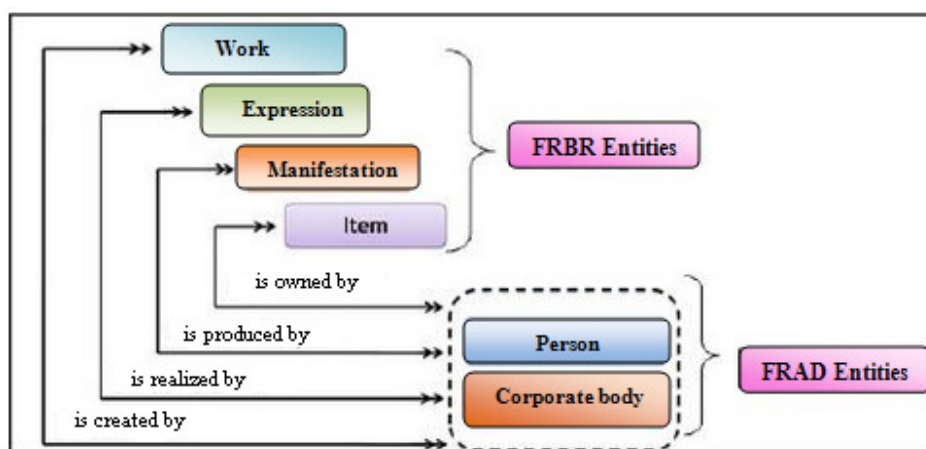
The new FRBR (Functional Requirements for Bibliographic Records) conceptual model was inspired by the Entity-Relationship (E-R) metamodel created in the 1970s by the computer scientist Peter Chen to optimize database routines so that records generated in databases were Interconnected, speeding up the retrieval of important items easily and quickly. According to Moreno (2006, p.34):

Data models are a set of concepts used to describe a database. The E-R model is an object-based logical model, and the identification of entities and relationships is understood as the capture of data semantics to design a database. This model does not aim at the implementation, but at the modeling/representation of data. From this model, it is possible to implement a database in other data models: object-oriented, relational, etc.

According to Silva and Santos (2012) conceptual modeling describes that a domain is composed of things represented by entities, which in turn have characteristics to be described, named attributes; and by the relationships between them. That is why relational databases have greater semantic capacity.

The figure below illustrates the E-R model in a database with information about a set of objects classified as a person (FRAD entity), a set of objects classified as manifestation (FRBR entity), a set of associations, each one linking a manifestation to a person - created by (relationship) and a set of information that characterize entities (attributes).

Figure 1 - E-R Model (bibliographic domain)



Source: Assumpção (2012). Adapted by the authors

The model (E-R) was designed to improve the development of databases by providing more accurate searches for users. The FRBR conceptual model was also created by IFLA (International Federation of Library Associations and Institutions) to represent the needs of the users of catalogs and bibliographies taking into account their tasks or functions - find, identify, select and obtain - with the objective of approaching the to the work.

The FRAD (Functional Requirements for Authority Data) model is an extension of FRBR for authority registration, it corresponds to the authorized heading for an entity established according to the rules of a cataloging agency. In the bibliographic universe, a FRAD entity lists information about a person or entity, which has its name used as a controlled access point in the bibliographic records, either in a catalog or in a repository.

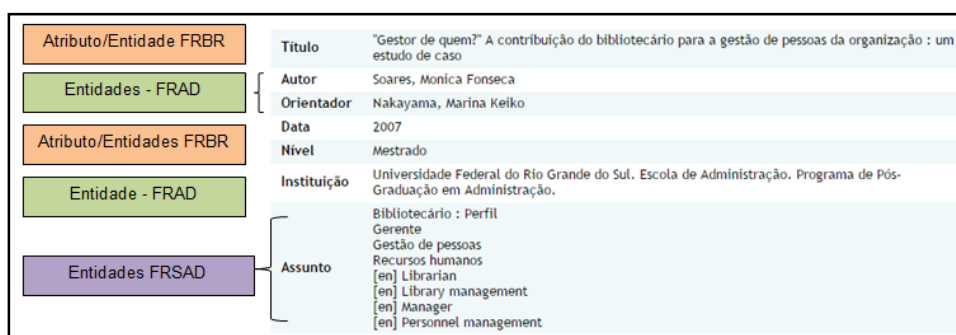
FRAD entities are responsible for the creation, production, editing and other responsibilities related to a work. Its use extends not only the capacity of relationships such as data associations, since when individualizing an entity by the application of attributes, we can associate it with open data, such as lattes curriculum, research groups, etc. mainly for academic authors with production in digital repositories.

Application of the FRAD model in a repository

The use of the DC standard associated to the E-R model will enable the user/author when depositing a document to describe it more precisely, since the new RDA (Resource Description and Access) standard allows the description of bibliographic and authority metadata in an extensive way, in language close to the one used by users when defining entities and their attributes. ²⁸³

To exemplify, we use the institutional repository LUME of the Federal University of Rio Grande do Sul as an empirical field. To begin the process of modeling this repository, we contextualize its structure, approaching it from the entities of the FRBR family. For the purpose of the study, we will model the FRAD entities, adding attributes that identify and qualify the people responsible for the production of the work (a master's dissertation) and the related research entity/institution.

Figure 2 - Identification of entities



Source: by the authors

The simulation below demonstrates the modeling of the descriptive structure of the Institutional Repository (RI) LUME, fields referring to authority data. The addition of FRAD attributes to DC elements contributor.advisor and contributor.author that represent access points/person of the work " *Gestor de quem?*" *A contribuição do bibliotecário para a gestão de pessoas da organização : um estudo de caso*" aims to improve the process of information search and retrieval in the sense that allows the repository users the new possibilities of relationships during the research. With this, the user will have more information about the creator of the document/resource and the different relationships it

maintains with other entities, in the case of our example, its supervisor, the program to which it is affiliated, research groups, etc.

In this direction, we can say that the addition of new attributes (lattes curriculum, institutional affiliations, etc.) to authority metadata identified in the DC record (author, advisor) expands the information search and retrieval processes; lead the researcher to navigate beyond the content of the work as it contextualized it in relation to its peers. For resources such as dissertations and theses, such relationships are fundamental for the dialogues among researchers, which makes the repositories a powerful scientific communication tool.

In LUME's repository case, it would be interesting to describe the metadata of the board members so that the relations are more complete. However, this information was suppressed because the manager, the Federal University of Rio Grande do Sul, follows the structural metadata (MTD-BR) defined for this type of work (theses and dissertations) defined by BDTD (Brazilian Digital Library of Theses and Dissertations), since they are part of this consortium coordinated by IBICT.

The figure below shows the application of FRAD conceptual model in the descriptive structure of RI-LUME referring to authority data (person entity) with the addition of attributes that identify author and advisor authorities in a contextualized way, with respect to their academic insertion. In addition to the extensible form, we also use headings for entities related to names, as this is the normalized form prevalent in most catalogs, which allows us to interconnect data about different scientific communities.

Figure 3 - Application of FRAD in RI-LUME

Entidade FRAD	dc.contributor.advisor	Nakayama, Marina Keiko
Atributo FRAD	Link Curiculo Lattes	http://buscatextual.cnpq.br/buscatextual/visualizacv.do?id=K4798728D2
Atributo FRAD	Linha de pesquisa	Gestão do conhecimento e compartilhamento
Atributo FRAD	Grupo de pesquisa	Núcleo Estudos e Observação de Gestão, Aprendizagem e Pessoas
Atributo FRAD	Link grupo/pesquisa	http://www.neogap.ufsc.br/wp/?page_id=58
Entidade FRAD	dc.contributor.author	Soares, Monica Fonseca
Atributo FRAD	Link Curriculo Lattes	http://buscatextual.cnpq.br/buscatextual/visualizacv.do?id=K4782635D0
Atributo FRAD	Programa/Pós-graduação	Programa de Pós-Graduação em Administração
Atributo FRAD	Forma normalizada do nome do Programa	Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Administração
Atributo FRAD	Link para o programa	http://www.ufrgs.br/escoladeadministracao/pos-graduacao/

Source: by the authors

In the institutionalization of a scientific field, reliability in the knowledge produced is necessary. An institutional repository is a means of disseminating the knowledge and results derived from the research undertaken by scientists; a communication channel that sheds different formal and informal productions whose results can be accessed and shared by other researchers. Therefore, in addition to the flexibility provided by the application of the FRBR model, it is up to the manager to be accurate in describing and coding the metadata, taking into account the use of RDA and RDF language, thus assessing the advantages arising from the use of these standards in the information retrieval and access in digital environments. ²⁸⁵

Partial conclusions

As we have seen, an institutional repository has the objective of preserving knowledge accumulated in institutions, presenting as main advantages the storage of a large amount of information, ease of search and retrieval and autonomy for the producer to make the description and self-archiving its documents.

The simulation process of the application of FRAD attributes in descriptive structuring has demonstrated that such implementation will considerably improve information search and retrieval process, qualifying the records with the addition of individual attributes such asattes curriculum and academic contextualization, such as linkage to groups and/or research centers, which will allow the user a more comprehensive view of the scientific communities in which the authors and their peers are inserted.

Regarding information retrieval, two aspects deserve to be highlighted: a) the detailing of a record with attributes additions extends the possibilities of relationships between entities and other bibliographic entities that compose the DC record; b) incorporating links (URL), such as theattes curriculum, opens the possibility of associating bibliographic tools with open data available on the web.

References

The references was made following the ABNT rules.

ASSUMPÇÃO, Fabrício. *O que é FRBR?* Available in: <<http://fabricioassumpcao.com/2012/07/o-que-e-frbr.html>>. Accessed in 4 maio 2015.

CAMPOS, M. Luiza Machado; CAMPOS, M. Luiza de Almeida; CAMPOS, Linair Maria. Web semântica e a gestão de conteúdos informacionais. In: MARCONDES, Carlos Henrique et al. (Org.). *Bibliotecas Digitais: saberes e práticas*. Salvador, EDUFBA; Brasília; IBICT, 2005. p. 55-286 74.

DCMI. *Using Dublin Core*. Disponível em: <http://dublincore.org/documents/usageguide/> Acesso 30 abr. 2015.

MARCONDES, Carlos Henrique. Metadados: descrição e recuperação na Web In: _____ et al. (Org.). *Bibliotecas digitais: saberes e práticas*. Salvador, BA: EDUFBA; Brasília; IBICT, 2005. p. 77-143

MORENO, Fernanda Passini. *Requisitos funcionais para registros bibliográficos – FRBR: um estudo no catálogo da Rede Bibliodata*. 2006. 199 f. Dissertação (Mestrado em Ciência da Informação) – Universidade de Brasília, 2006.

SILVA, Renata Eleutério da; SANTOS, Plácida Leopoldina Ventura Amorim da Costa. Requisitos Funcionais para Registros Bibliográficos (FRBR): considerações sobre o modelo e sua implementabilidade. *RBBD*. São Paulo, v. 8, n. 2, p.116-129, Jul./Dez. 2012.

SOUZA, M. I. F.; VENDRUSCULO, L. G.; MELO, G. C. Metadados para a descrição de recursos de informação eletrônica: utilização do padrão Dublin Core. *Ciência da Informação*, v. 29, n. 1, p. 93-102, Jan./Abr. 2000.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL. *Repositório Lume*. 2015. Disponível em: <http://www.lume.ufrgs.br> Acesso 30 abr. 2015.

Nanopublishing Modeling: an experimental approach



Lorena Tavares de Paula
Federal University of Minas Gerais
lorena.ltp@gmail.com



Maria Aparecida Moura
Federal University of Minas Gerais
cidamoura@gmail.com

Introduction

Information organization in digital environments requires the amplification of theoretical foundations and methodologies for the treatment of documentary mass. Therefore, there is a need to develop proposals for organizing information that guide the professional practice considering the new scenario of documentary production.

In this sense, the present article highlights a methodological experiment for the modeling of nanopublishings in interface with indexing in digital environments.

Nanopublishing indexing: conceptual aspects

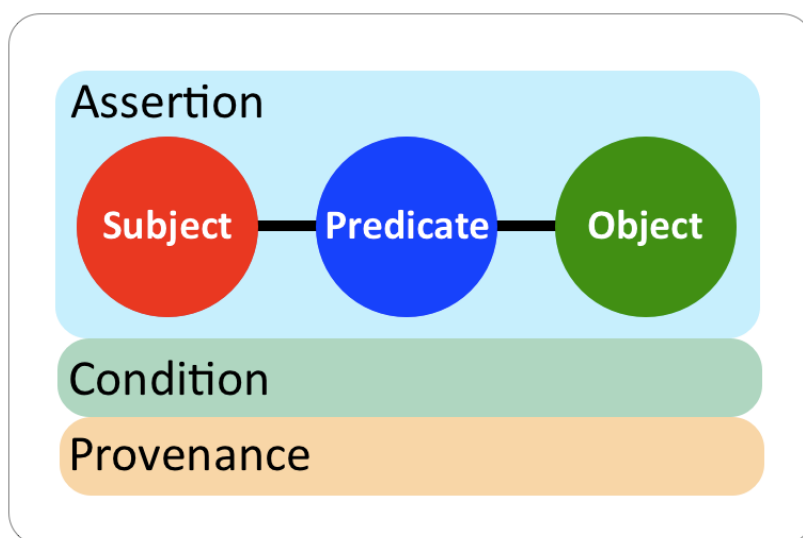
The proposal of elaborating nanopublishing, established by the Concept Web Alliance (CWA), aims at the development of systems that delimit standards for the representation of scientific information. In a simplified description, it is possible to state that nanopublishings are effected through the mapping of concepts arranged in digital publications in a given knowledge domain.

Nanopublishing enables communication that is based on high quality contextual information and supports the dissemination, appropriation and organization of contextualized information according to an author and a conceptual unit. Groth, Gibson and Velterop (2011) explain that the elaboration of a nanopublishing requires structural

contents that reinforce the value of a declaration to the point where it can be considered a publication.

Nanopublishing is, therefore, a computational application for the representation of conceptual assertions in context.

Figure 1: Nanopublishing Anatomy



Source: nanopub.org

According to FIGURE 1, the "nanopublishing anatomy" is characterized by an Assertion signaled by the semantic relation established by "subject, predicate and object". It also has statements (condition) conditioned to the identification of the assertion. In addition, it is characterized by provenances (Provenance) that establish the identification of the elements, which legitimize the nanopublishing, such as: author, location (URL), institutional links, among others. The elements "assertion"; "condition" and "provenance" are embodied in computational applications that use open standards, such as XML (Extensible Markup Language), and RDF (Resource Description Framework) to foster collaborative data integration.

In its conceptual aspect, nanopublishings should be citable, attributable and reviewable. In addition, they need to be reorganizable as they must be easily aggregated and identified across the web. According to Groth, Gibson and Velterop (2011), they need to be extensible to meet the new forms of description oriented by sets of definitions.

Thus, nanopublishings establish a thematic representation of the document, through conceptual evidence and the affirmation of authorship and knowledge communities.

In the field of information organization, the indexing process aims to establish the subject (aboutness) of a document obeying two basic principles: | 289

- 1 - Conceptual Analysis;
- 2 - Translation to an indexing language.

Conceptual analysis involves deciding what the document is about. An efficient subject indexing suggests that a decision is made not only as to what is dealt with in the document, but also as to why it is likely to be thematically interested in a given group of users. The translation phase for an indexing language is related to the need for vocabulary control, in order to improve information retrieval process.

It is worth mentioning that the entire indexing process aims to improve the specific representation of a document and its consequent retrieval, which Cesarino (1985, p. 85) describes as a "set of consecutive cooperations executed to locate, within the totality of available information, that really relevant". From the description of the subject in the form of descriptor terms, it is possible to promote an integration among document, collection and user.

Nanopublishing aims, therefore, to provide a representation of contents, through the structuring of conceptual statements contained in a formal publication. Therefore, an experimental methodology is presented that aims to integrate the indexing process into the modeling of nanopublishings derived from electronic documents.

Methodology for nanopublishing modeling

The methodology for a nanopublishing modeling incorporates the principles of indexing, as well as three dimensions essential to information and knowledge representation in digital contexts: the informational dimension, the computational dimension and the interface.

In scientific methods, an experiment is a set of actions regarding physical phenomena. The experiment is based on an empirical approach. Depending on the

perspective, it may lead to an objective and complete understanding of the physical world or will only lead to improvement of subjective knowledge through theoretical/methodological apprehensions.

According to Gil (2002, p. 69), the experiment can be considered a hypothetical-deductive method, so that, for the author, "the scientist, through a combination of careful observation and skillful anticipations, reaches a set of postulates that govern the phenomena he is interested." Experiments are, therefore, highly accepted, especially in natural sciences. Gil (2002, p. 73) also states that "in neopositivist circles, it is even considered as the only rigorously logical method." However, the author also points out that "its application to social sciences presents serious limitations". "Nevertheless, many authors consider it sufficient for the construction of logical model of investigation in social sciences".

Regarding Information Science, it is believed that, as an applied social science, it can offer to the field methods that conciliate the integration of processes for the accomplishment of experimentation as a research method. Although the social context may suffer adaptations and changes, a documentary corpus selected for an experimentation modeling will suffer little from changes in the sociocultural context. In this sense, the experiments proposed from this methodological experiment can be understood as a simulation.

A simulation is a "virtual experiment" that requires an operative model representing a system or processes that characterize it in the whole (or in part). The use of this methodology is analyzed as a way to search for a model, to confirm it or to make a projection of future events. In this way, the simulation lends itself both to the "context of discovery" and to the "context of the proof".(VICENTE, 1995, on-line)

The methodology to develop nanopublishing modeling experiments considers the hypothesis that nanopublishing can be a model of information representation in a digital environment, when placed on interface with indexing principles.

Regarding these considerations, the phases of the methodology, the modeling experiment, are described in the following sections.

Phase 1 - Generation of the mapping of the corpus to be researched: informational dimension

In a first step, it is necessary to select journals from correlative fields of knowledge and that use and disseminate the concept to be modeled in nanopublishing format. For all entries of a given concept, it is possible to extract the following metadata: article title, journal title, volume, number, article homepage, publication date, author names and surnames, URL.

Table 1 - Identification elements of "phase 1" corpus (nanopublishing provenance)

Concept	
Article title	
Journal title	
Number, volume	
Initial page of article that carries the concept	
Publication date	
Name of the author	
URI - (<i>Uniform Resource Identifier</i>).	

Source: by the authors

TABLE 1 presents the metadata to establish a structure that allows to consult, identify and locate the conceptual information in the body of the articles. The plan for extracting and identifying the concepts that could make up the nanopublishing will depend directly on the information consolidated from the table described in phase 1. This information will materialize the e-evidence of nanopublishings.

As an example, one can consider the intention of nanopublishing modeling for the concept "innovation". The first step would be to search for the term "innovation" in a database, such as Scielo, in which results of this concept would be obtained in articles from journals of different fields of knowledge. The corpus to be formed to represent the

modeling of the concept "innovation" should be described following the guidelines of the table described in phase 1.

3.2 Phase 2 - Conceptual extraction from indexing principles: informational dimension

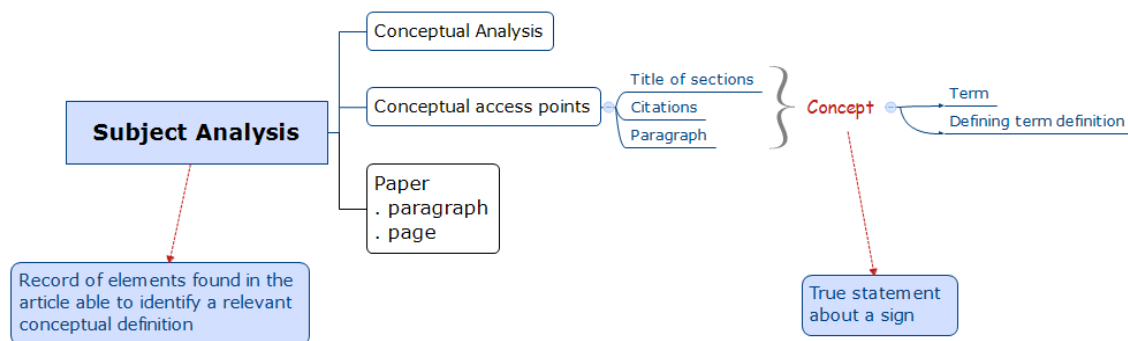
The concepts to be modeled in nanopublishing can be general or individual and reveal the complexity of the domain they are associated.

HJORLAND (2001, p. 774), citing Wilson (1968), highlights the most important elements in determining the subject of a document for extracting relevant concepts:

- a) identifying the author's purpose in writing the document;
- b) identifying the domain and subordination of the different conceptual elements;
- c) observing a set of concepts and established references;
- d) establishing a set of selection rules for the "essential" (as opposed to non-essential) elements of the document as a whole.

For Hjørland (2001, p. 774), the method for determining the subject of a document is closely related to the theories of meaning, interpretation, and epistemology. In this sense, the process of conceptual extraction proposed in this methodology follows the perspective of the identification of subject proposed by Hjørland.

Figure 2 - Conceptual extraction



Source: by the author, 2015.

According to FIGURE 2, the subject analysis of each article that composes the corpus is carried out. In this process, the concepts that have explicit definitions and help in the identification of the evidenced knowledge domain are selected. These concepts can

be extracted from: section titles, citations, footnotes and paragraphs. Subsequently, the concept is identified through a "term" and a "statement" that establishes the nominal definition.

Following the example with the word "innovation", the fundamental concepts that are part of the formulation of the definition "innovation" will be identified through the subject analysis, always observing the different areas of knowledge in which the concept is formulated. ²⁹³

In this phase, marking, in the body of the article, the page and paragraph that highlight the concept is essential. Following are the proper authorship references:

- a) authors of the registry;
- b) authors cited for the conceptual formulation;
- c) origin knowledge community (which field of knowledge the author of the record and the cited author belong to).
- d) statements for the conceptual definition

From the identification, characterization and marking of the concepts, along with their definitions and characterization of the knowledge domain, one can begin the nanopublishing modeling

3.3 Phase 3 - Nanopublishing Modeling: informational dimension

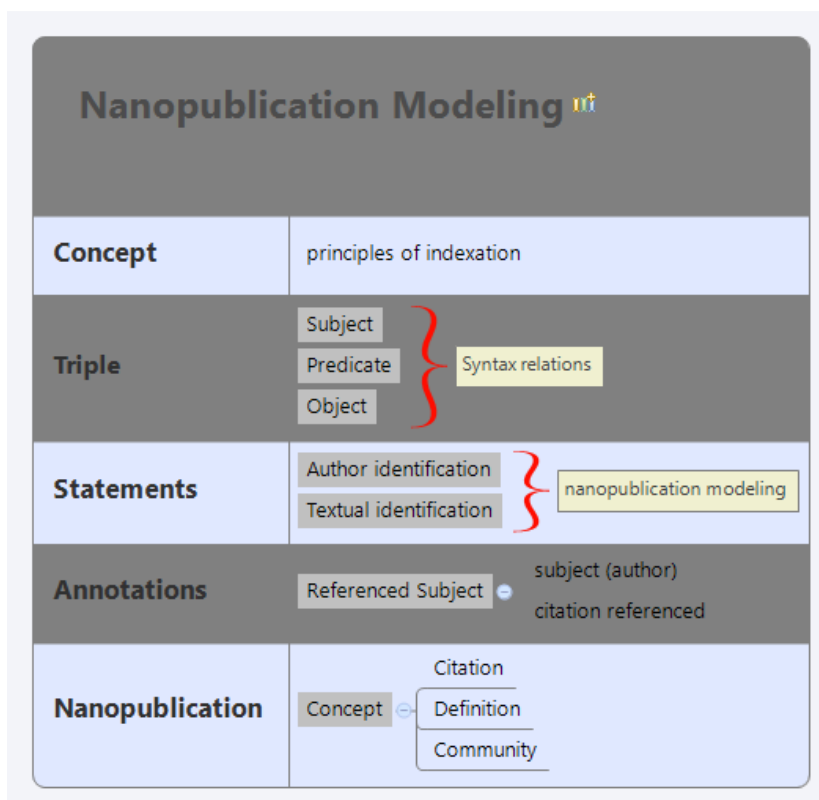
Following the definitions of phase 1, which establishes the identifiers of the selected corpus, and phase 2 with the process of identification and conceptual selection, it is expected to establish identifications, such as:

- a) identifications of conceptual definitions (phase 2)
- b) representative authors (phase 2)
- c) fields of knowledge in which authors and concepts act (phase 1)
- d) location of publications (phase 1)

From the identification of these elements, one can start the nanopublishing modeling following the perspectives of the triple syntax: subject - predicate - object.

Following the example with the concept of "innovation", syntax relations will be established in the following order: Subject = Innovation / predicate = link connectors (verb) / Object: conceptual formulation.

Figure 3 - Nanopublishing Modeling



In FIGURE 3, we observe that the first action consists in the identification of the concept. Proper extraction of conceptual definition is essential to delimit understanding and reuse of nanopublishings. In the definition, the provenance must be highlighted, according to the processes delimited in phase 1 (corpus description).

In sequence, we define the triple syntax processes: subject - predicate - object. At this point, nanopublishing is described in its conceptual relations for the construction of assertions that will support the statement.

The assertions consist of a textual identification related to an authorial identification. Nanopublishing is described with an approach oriented by structured data representation, along with its provenance in a publishable and citable entity.

Thus, nanopublishing consists of the assertion, verified in the triple syntax, and the provenance of the assertion is materialized in the contextualized statement with authorships and communities of knowledge that deal with the concept.

Annotations offer the completeness of the subjects mentioned by nanopublishing. These subjects include the identification of the record author and the cited author in the references of the article, as well as the assigned conceptual definitions. ²⁹⁵

It should be noted that these elements are the embodiment of nanopublishing and present mechanisms to guarantee the integrity of the concept and its dissemination origin together with the author responsibility. In addition, they guarantee access to provenance, allowing users to assess the reliability of concepts. Above all, they provide a mechanism by which authors and institutions can be recognized for their contribution to formulating a concept in a given scientific field.

Thus, the listed publications will be the identification URL of the nanopublishing that will offer elements for the computational application of the nanopublishing. The combination of assertion and provenance of the assertion, which materializes the origin of the nanopublishing, represents the publishable and citable entity.

3.4 Phase 4 - Technological aspects - Semantic Web application in nanopublishing modeling: computational dimension

Nanopublishing can expose quantitative and qualitative data.

The computational representation of Nanopublishings can be consulted in an automated way using Semantic Web technologies. The computational application oriented to nanopublishing is based on RDF (Resource Description Framework). Representations of the triple syntax (Subject, predicate, object), characteristic of nanopublishing, can be established from RDF processing. This technology suggests a representation for the web content, which allows the integration of databases and analysis of information by software applications, based on Uniform Resource Identifier (URI).

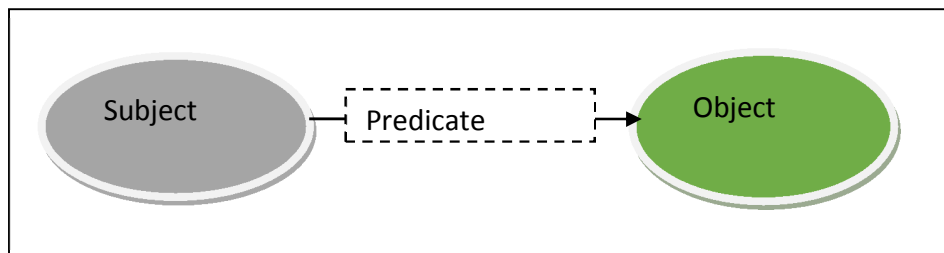
Nanopublishing description is made essentially by assertions or statements that, graphically, denote the connection between two nodes of a graph. From this perspective, the first node is the Subject, while the destination node is called Object. The segment that

connects subject and object is the predicate. In this model all elements are tagged by the respective URI.

The following figure shows an example of an RDF graph with only one assertion.

Figure 4 - Simplified RDF mode

Source: by the authors



The nanopublishing modeling in RDF may follow criteria suggested by nanopb.org (2015):

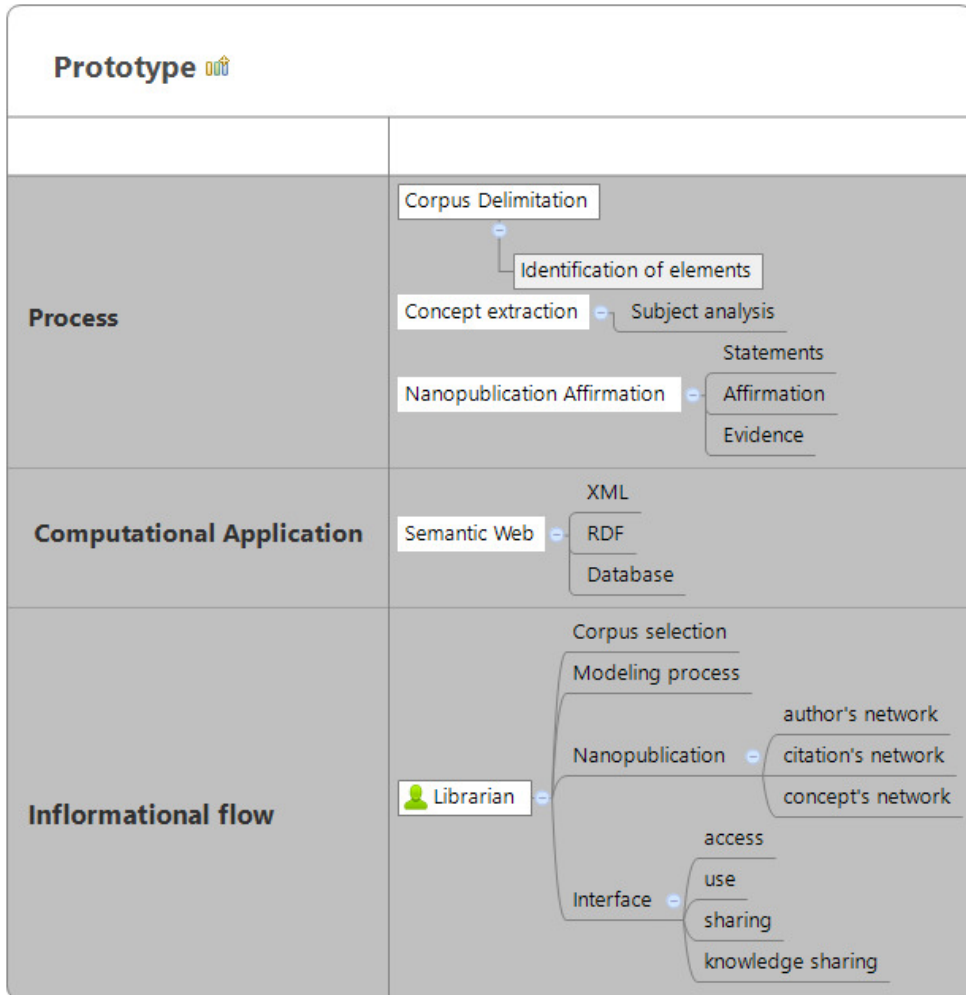
- a) the publication will be guided by a set of four RDF (subject-predicate-object + context);
- b) the context (the graph) of each triple must be specified as a valid URI;
- c) explanations of Declarations;
- d) explanations of Assertions;
- e) materialization of E-evidence of nanopublishing.

The use of RDF to represent the nanopublishing modeling is justified in its simple descriptive processes. In addition, RDF makes it possible to present and explore cooperative work (MANOLA; MILLER, 2002). This happens through the authentication of several namespaces relating concepts and authorships in communities of different contexts.

Phase 5 - Interface Proposition

An interface based on semantic Web technologies must follow guidelines that help the user understand the overall properties of the system. In this way, it is necessary to propose essential elements that characterize the functionalities intended in the system for information organization. In this conception, one can suggest a priori the elements of information organization that are fundamental to the system:

Figure 5 - Interface suggestion



According to FIGURE 5, the "Input System" should be based on the corpus identification processes and indexing principles that are presented in phases 1 (one) and 2 (two) of the methodology. All input elements, in which the selected corpus is mapped for further nanopublishing modeling, can be controlled by the manager.

The nanopublishing indexer establishes the processes of conceptual extraction and the applications of the triple syntax in the elaboration of the statements and assertions of the nanopublishing.

It should be noted, however, that in indexing-oriented processes, not all aspects of nanopublishing can be established by automated processes of semantic web technologies.

Next, one must apply the semantic web-based technological layer. This computational layer will provide the necessary technical adjustments to the automation,²⁹⁸ management, and dissemination processes of nanopublishings in line with interoperability. Finally, in the exit system, it reveals a network of information based on the declarations, assertions and e-evidences of the nanopublishing.

In the computational application interface, in turn, there is a need to use XML and RDF. These applications must provide a computational infrastructure that contains both a relational database and the possibility of collaborative information construction. These technological functions should promote, at the exit of the system, the consolidation of the nanopublishing declaration and the identification guidelines in the form of an authorial and institutional network woven from the elements of e-evidence.

Finally, it should be pointed out that the information flow observed in the interface system demonstrates the role of the manager in his/her function of separating the corpus and applying the essential methodological phases to the experiments. From this action, one has the perspective of establishing nanopublishing and their authorship networks, concepts and citations.

Conclusion

In classic indexing procedures, information representation occurs with terms that correspond simultaneously to the intellectual content of the document and to the probable query from the user's search. However, in view of the growing volume of information published in digital media, it may be necessary to select contextualized "conceptual units" that help in the construction of new knowledge. It is suggested that such units can be considered "nanodocuments". These, in turn, are the smallest cited conceptual unit free from ambiguities and are based on the assertion of an author held in formal publications.

In this way, the hypothesis is that nanopublishing can be considered an emerging model of information representation in digital environments with methodological orientations established in dialogue with indexing. The confirmation of this can be

obtained through experiments and simulations that have, in the methodology suggested in this article, the fundamental guiding elements.

In digital environments, therefore, there is an urgent demand for the creation of mechanisms of information representation that are able to specify an informational content, without distortions in relation to context and conceptual understanding. In this respect, the integration of indexing principles with the nanopublishing modeling process can be considered a promising methodology.

References

The references was made following the ABNT rules.

CESARINO, M. A. N.; PINTO, M. C. M. F. Análise de assunto. *Revista de Biblioteconomia de Brasília*, Brasília, v. 8, n. 1, p. 32-43, jan./jun. 1985.

DAHLBERG, Ingetraut. Teoria do conceito. *Ciência da Informação*, Brasília, v. 7, n.2, p. 101-107, 1978.

GIL, Antônio Carlos. *Como elaborar projetos de pesquisa*. 4. ed. São Paulo: Atlas, 2002. 175 p.

GOMES, Hagar Espanha; GUSMÃO, Heloísa Rios. *Guia prático para elaboração de índice*. Niterói: APB-RJ, 1983.

GROTH, P; GIBSON.A; VELTEROP, J. *The Anatomy of a Nano-publication*. Disponível em <<http://www.w3.org/wiki/images/c/c0/hclsigswansiocactionsrhetoricalstructuremeetings20100215cwa-anatomy-nanopub-v3.pdf>> Acesso em: 14 mar. 2014.

HJORLAND, Birger. Information Seeking and Subject Representation. *An Activity-theoretical approach to Information Science*. Westport & London: Greenwood Press, 2001.

KLEIBOER, M. Simulation methodology for crisis management support. *Journal of Contingencies and Crisis Management*, Nijmegen, v. 5, n. 4, p. 198-206, Dec. 1997.

LANCASTER, F. W. *Indexação e resumo; teoria e prática*. Brasília: Briquet de Lemos/Livros, 2004. 347p.

MANOLA, F., MILLER, E. *RDF primer*. W3C Recommendation, 2004. Disponível em: <<http://www.w3.org/TR/rdf-primer/>>. Acesso em: 05 mai. 2015.

MOURA. M.A. Informação, Ferramentas ontológicas e Redes Sociais: a interoperabilidade na construção de Tesouros Ontológicos. In: Nuevas perspectivas para la difusión y organización del conocimiento, ISKO España 9, Valencia, Universidad Politécnica de Valencia, **Actas...**, Valencia, 2009.

MOURA, M.A. Folksonomias, Redes Sociais e a formação para o Tagging Literacy: desafios para

a organização da informação em ambientes colaborativos virtuais. *Inf Inf*, Londrina, v. 14, n.1, p.25-45, 2009.

MONS, B; VELTEROP, J. Nanopublication in the e-science Era. In: *Workshop on Semantic Web Applications in Scientific Discourse*, Washington, **Actas....**, 2009.

NANOPUB.ORG. Available at <<http://nanopub.org/wordpress/>>. Accessed in: 14 mar. 2014. .00

The Open PHACTS Nanopublication Guidelines. Available at: <<http://www.nanopub.org/guidelines/1.8/>>. Accessed in: 14 mar. 2014.

VICENTE, P. O uso de simulação como metodologia de pesquisa em ciências sociais. *Cadernos EBAPE.BR*, v. 3, n. 1, mar. 2005.

WILSON, Patrick. *Two Kinds of Power: An Essay on Bibliographical Control*. Berkeley: University of California Press, 1968.

Application of ontologies in the information retrieval process applied in academic settings



Caio Saraiva Coneglian
São Paulo State University
caio.coneglian@gmail.com



Elvis Fusco
University Euripides of Marília
fusco@univem.edu.br



José Eduardo Santarém Segundo
University of São Paulo
santarem@usp.br

Introduction

The mass data explosion is testing the capabilities of the most advanced storage, processing, transformation, and information analysis technologies. The areas of information processing and retrieval are being challenged by the volume, variety and speed of a flood of semi-structured and unstructured data of complex nature.

Therefore, there is a large amount of data spread over the Web, presenting great difficulty in organizing all this information and thus, generate knowledge to individuals. Bastos (2005) claims that the generation of knowledge only occurs with the assimilation of organized information, because it is able to establish relations of meanings with knowledge.

Information Science, as Borko (1968) discusses, is a discipline that covers information origin, collection, organization, storage, retrieval, interpretation, transmission, transformation and use. In this way the areas of information retrieval and knowledge organization are related, so that one is able to complement the other in their objects of study. | 302

Thus, in order to deal with the issue of how to organize information and thus to present more efficient results, semantic concepts are associated with the Information Retrieval process, with the purpose of adding meaning to the contents sought in a specific domain, which allow the search no longer by key words in a textual search process, but by meaning and value, extracting from Web pages and services the information of real relevance, discarding what is unnecessary. From this, ontology appears as solution in the search to insert semantics in this process.

This research aims to create a semantic platform for Information Retrieval on the Web that allows location, storage, treatment, organization and retrieval of information inserted in the Web context in the most varied information sources on the Internet that serve as the basis for a computational architecture that transform the disaggregated information into a strategic, relevant, accurate and usable information environment to allow users to access information with greater added value, that is, information able to satisfy the user's informational needs, attaching semantics to the Information Retrieval process.

Methodology

The present study is characterized as a descriptive and analytical research, based on documentary analysis, with an exploratory characteristic, focusing on the implementation of a semantic retrieval platform of information applied to the Web. The construction of the project was divided into three main parts, bibliographic search, ontology construction and prototype implementation.

Information retrieval

Traditional information systems are unable to efficiently cope with all the new dynamic data sources and multiple information contexts that mainly have the Internet as a platform.

Problems are encountered in retrieving, standardizing, storing, processing and using information generated by a number of heterogeneous sources that serve as a basis for nurturing the decision support systems of organizations.

In this context, it is questioned whether the information computational environments really present the information that is relevant to decision makers in organizations.

For this, the creation of an Information Retrieval architecture in the Web context was proposed, as seen in Figure 1.

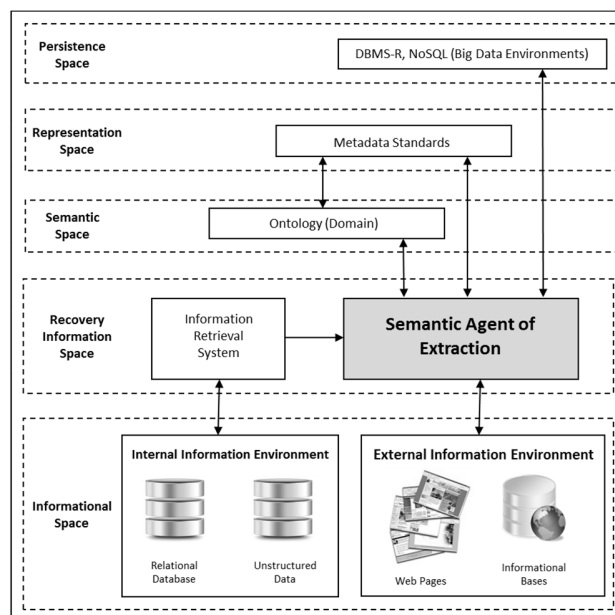


Figure 1: Contextualization of the Semantic Agent Architecture

This architecture was proposed so that information retrieval can be carried out using the semantic space. The architecture is centered on the information retrieval agent and integrates with all other elements and layers.

This architecture will be used in a higher education institution as in this area information is not being used efficiently in competitive intelligence.

This architecture is distinguished from others, as it will analyze all the retrieved information, verifying whether the information is contained inside a domain, therefore, only information related to the problem that the user has will be extracted, according to the defined ontology.

Ontology

The word ontology comes from *ontos* (being, entity) and *logos* (knowing, doctrine), and strictly means the "study of being". It arose from the study of philosophers, still in Aristotle's time, and was used in this context to make an approach to being as a being, or to being in general. Later still in philosophy, the term ontology became more used to know what is fundamental or irreducible, common to all beings.

Ontology is defined as a logical theory that represents an intended vocabulary, that is, it is a contextualization of something particular existing in the world. In this sense it is observed that with an ontology one can define particular contexts and domains of the world (GUARINO, 1998)

Gruber (1995) claims that in a context of multiple agents, ontology could define the context, the vocabulary of that domain, thus serving as the basis for communication between agents, and to be able to make their extractions in the knowledge in which they are present. Gruber (1995) adds that ontology is an explicit specification of a conceptualization.

According to Santarém Segundo (2010), Information Science and Computer Science used ontology when referring to the acquisition of knowledge from semi-structured data, using ontology to apply techniques and methods, to process information.

Santarém Segundo (2010) claims that ontologies come with the main objective of having shared vocabulary, where information can be exchanged, and used by other users. These users are both human beings and intelligent agents.

For ontology construction, it is necessary to use a methodology that allows the built ontology to be consistent. For that, the methodology proposed by Noy and McGuinness (2001a) was used, in which it is explained that for the ontology construction seven steps must be followed.

Semantic agent of extraction

The creation of a software agent that semantically aggregates the information available on the Web of a particular domain can bring subsidies to a computational platform for the creation of an informational decision support environment that gives a broader view of the internal and external information scenarios of relevance in organizational management. | 305

This agent should communicate with the internal and external information spaces of the Web basing their searches on ontological rules based on a metadata standard to perform the semantic extraction of the proposed domain and will support other systems in a broader context of Information Retrieval.

The research uses the domain of higher education institutions as a case study to apply the computational platform proposed in the architecture described in Figure 1. As ontology prototype, a course on database was used to search for scientific articles.

Ontology Conceptual Notation

The first step in creating the ontology was to check what terms of the course database are.

After verifying these terms, the hierarchy was structured between the classes of the ontology, this was realized according to the experience of the authors and of other researchers in the area.

Ontology conceptual notation was elaborated using Protege software (NOY, 2001b), as shown in Figure 3, and the relationship between classes was also constructed.

The agent will use this ontology to perform a task, so this ontology that was constructed is classified as a task ontology, according to Mizoguchi (2003) classification.

A task ontology is an ontology that solves a specific problem within a domain, that is, it solves the problem of scientific research in the domain of course database.

After implementing the ontology in Protégé software, the class diagram with its properties was created, using OWL (Web Ontology Language). After the creation of the

OWL file, Owl2Java software (ZIMMERMANN, 2009) was used to transform OWL into Java classes; thus making the implemented ontology.

Functioning of the Semantic Agent

The search agent captures information through pre-defined web pages and uses ontology to classify and after that perform a search semantically.

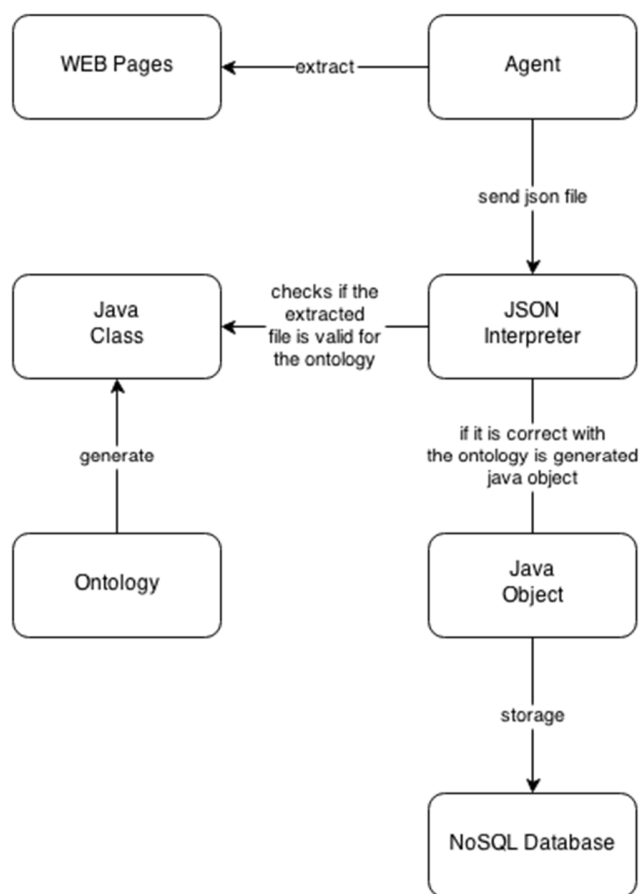


Figure 2: Agent functioning

Figure 2 shows the process performed by the system. The user performs a search on any subject related to the area of database, the agent extracts abstracts from the databases related to this topic. These abstracts should be analyzed in a process where they take into account the words contained in this abstract that are present in the topic

searched by the user. In this way it is possible to use a built ontology, which deals with a specific topic in the area of scientific research, in this case, in the field of the course on database.

The sequence of tasks performed by the agent is described as follows:

Information Extraction

The agent extracts from the IEEE Xplore page (<http://ieeexplore.ieee.org>), the abstracts, based on the research the user performs. Based on the location of the abstracts on the HTML on the page, the agent extracts information, and transforms it into a String.

The agent process is divided into three phases: search the page, extraction of titles and abstracts and return a list of articles to the main program.

- Search the HTML page: this first phase is characterized by performing a search in the search engine of IEEE Xplore, and extracting the HTML that the page returns as a result of this search.

- Extracting titles and page abstracts: after the return of HTML, the agent extracts from this, title and abstract of each article. This is possible through an analysis of the HTML page, checking the tags whose data from abstracts and titles are inserted.

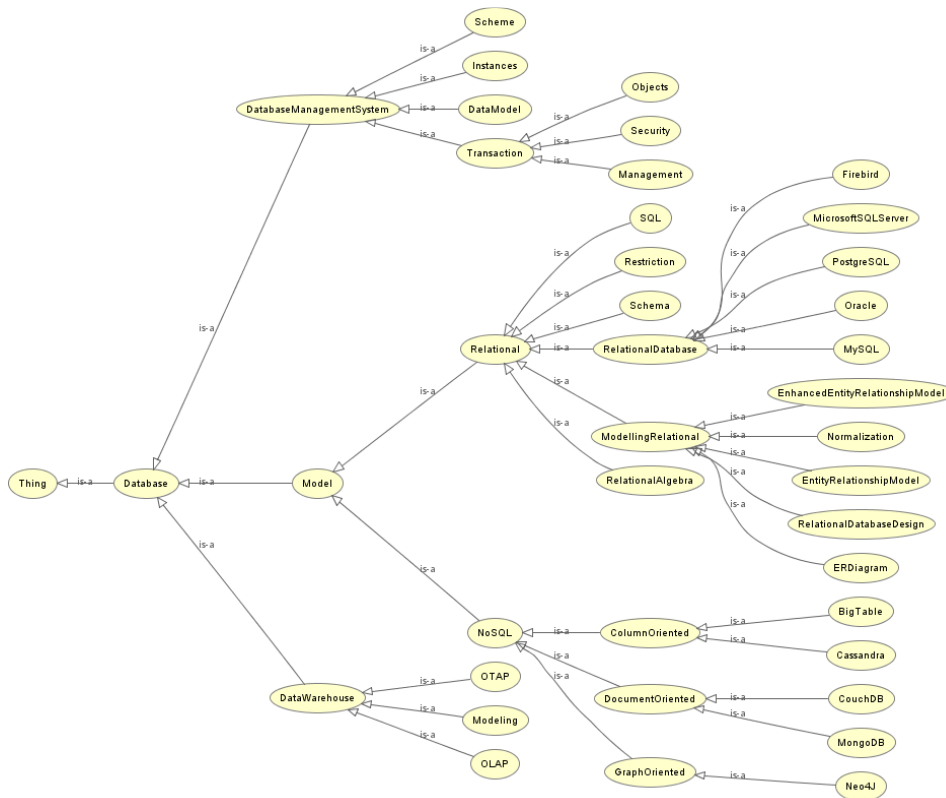


Figure 3: Class relation of the ontology

In this way, this search robot is able to perform a syntactic extraction of the articles contained in the IEEE Xplore database, since the search robot retrieves the articles that were indexed by the database itself, creating a list with all the articles that were presented, to be used in the ontology.

Integrating Ontology with the Extraction Agent

In order for the program to actually have the presented semantics, the program uses the ontology to evaluate which of the results that have been extracted from the database, are really useful, and are related to the context of that search.

This integration happens in five moments:

- First, it is verified where the term searched by the user is within the ontology.
- After that, which are the classes hierarchically superior and inferior to the searched term are obtained.

- Subsequently, the terms that are part of that hierarchy of the searched term are checked within the abstract and title of the articles searched.
- Afterwards, a comparison is made between the number of terms that are in the hierarchy and those that are contained within the abstract and the title of that article. This results in a percentage of the number of terms that are in the hierarchy, which are within the abstract and title of that article.
- Finally, the user is presented with all articles that have reached a percentage above 35%.

Results

As a test to find out if the system is extracting and verifying the semantics of the extracted articles, a search was carried out with the user searching for the term "Datawarehouse".

The hierarchy of the term Datawarehouse are the terms: Database, Datawarehouse, OLAP, OTAP and modeling.

In the case of 25 articles, 7 were those that met the requirements, and these are presented to users. To visualize how the program does the analysis of abstracts and titles, below in figure 4, an article of the ones that met the requirements is presented.

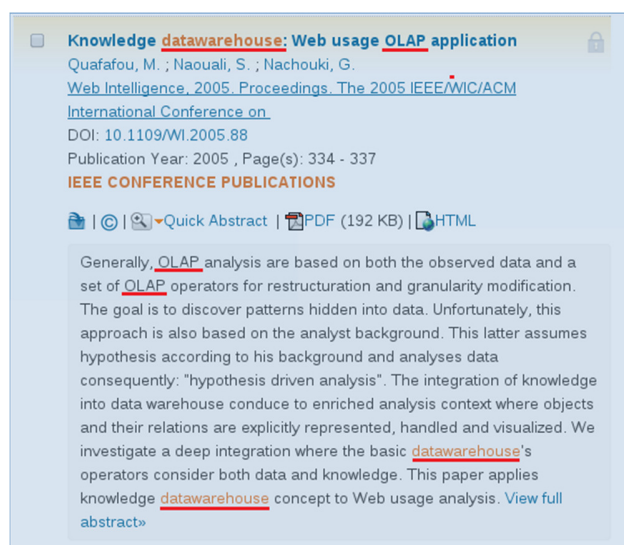


Figure 4: Example of an analyzed article

This paper presented 2 of the 5 terms of the ontology hierarchy. In this case it presented the terms OLAP and Datawarehouse. In figure 4 the terms that were found by the program are marked in red.

Conclusions

This paper presents the use of ontologies for improving the Information Retrieval process.

The purpose of this research is to adhere semantics to the process of Information Retrieval, using information within the context of Big Data, to carry out a process that adds more value to the searches performed by the user.

In order to prove this goal, the domain of scientific research was used, in which the user, when searching databases of scientific articles, faced with the problem of having a very large number of documents that do not meet the needs of the user.

It was then created an ontology and a search robot and made the connection between them to achieve the initial goal in this way.

After testing, it was observed that the use of ontology for the search agent is an effective way to obtain information of value and to meet the informational needs of the user.

Ontology can be efficient in the present process because it becomes a way of organizing semantic information, and thus, only meaningful information will be presented to the user.

Although the term Semantic Web has been used for a few years now, there is still a limitation on its use because much of the Web is organized in a syntactic way, in which most pages are created so that only human beings can read what is written there, without being structured in a way that computational agents can extract the data contained therein within a context, with an implicit meaning within the HTML.

Extracting agents are able to extract the documents from the Web and a program can, through the use of ontology, to treat the information, and thus to present the most relevant results to that user.

In this way, the results obtained with the use of the developed prototype, can refine the quantity of articles presented to the users. This research, therefore, tries to make the user obtain, in an Information Retrieval process, more expressive results and that presents greater value. This way, the user will be able to evaluate more expressive information, and will not waste time with data that do not meet their needs. | 311

Therefore, to address the issue of how to insert intelligence in retrieval of Web pages that do not present a contextualization of information, this research proposes that the process of adhering semantics to these pages occurs outside the Web, that is, the extraction of pages occurs in a syntactic way, and from what has been extracted, an analysis of the information occurs, inserting in this semantic way to this process. This method has proved to be very efficient, since it can actually perform a smarter search, which goes beyond simple search formulas, which only observe the syntax of the texts, and can analyze the context in which the extracted documents are inserted, and thus visualize whether that document meets the users' needs.

Acknowledgements

The presented paper was funded by FAPESP, process 2013/16369-3.

References

The references was made following the ABNT rules.

BASTOS, Flávia Maria. *Organização do conhecimento em bibliotecas digitais de Teses e Dissertações: análise da aplicabilidade das teorias macroestruturais para categorização de áreas de assunto*. 2005. 118 f. Dissertação (mestrado) - Universidade Estadual Paulista, Faculdade de Filosofia e Ciências, 2005.

BORKO, Harold. Information science: what is it? *American documentation*, v. 19, n. 1, p. 3-5, 1968.

BORST, Willem Nico. *Construction of engineering ontologies for knowledge sharing and reuse*. 1997. Tese (doutorado) - Universiteit Twente, 1997.

GRUBER, Thomas R. *Toward principles for the design of ontologies used for knowledge sharing?*. *International journal of human-computer studies*, v. 43, n. 5, p. 907-928, 1995.

FORMAL ONTOLOGY IN INFORMATION SYSTEMS: PROCEEDINGS OF THE FIRST INTERNATIONAL CONFERENCE, 1, 1998. Trento. *Proceedings...* Trento: IOS press.

MIZOGUCHI, Riichiro. Tutorial on ontological engineering. *New Generation Computing-Tokyo*, v. 21, n. 4, p. 363-364, 2003.

NOY, Natalya F.; McGUINNESS, Deborah L. *Ontology development 101: A guide to creating your first ontology*. Stanford, CA, 2001a. 25p.

NOY, Natalya F. et al. **Creating semantic web contents with protege-2000**. *IEEE intelligent systems*, v. 16, n. 2, p. 60-71, 2001b. | 312

SANTAREM SEGUNDO, José Eduardo. **Representação Iterativa. Um Modelo para Repositórios Digitais**. Marília, São Paulo, Brasil, 2010.

ZIMMERMANN, M. **Owl2Java-A Java Code Generator for OWL**, 2009.

Librarian performance in the subject analysis of theses within the theoretical dimensions of subject cataloging and indexing



Roberta Cristina Dal'Evedove Tartarotti
São Paulo State University
roberta_tartarotti@yahoo.com.br

Paula Regina Dal'Evedove
Federal University of Pernambuco
paula.dalevedove@ufpe.br



Mariângela Spotti Lopes Fujita
São Paulo State University
fujita@marilia.unesp.br

Introduction

In the scope of Knowledge Organization and Representation(KOR), Information Subject Treatment is highlighted as a fundamental subarea in coping with issues on "analysis, description and representation of document content, as well as their inevitable interfaces with information storage and retrieval theories and systems" (BARITÉ, 1997, p. 124).

Guimarães (2009, p. 106) clarifies that, in search of a theoretical foundation and establishment as a field of research, historically, Information Subject Treatment is presented under three theoretical currents: documentary analysis (*analyse documentaire*, from French conception); subject cataloging (from North American conception) and indexing (from English conception).

While subject cataloging is the dimension of cataloging process responsible for the subject analysis of informational items, as well as the provision of classification notations and subject headings that represent the subject matter of the items - turned

to online catalogs; indexing is a process formed by subprocesses/steps that aims to identify the contents of a document, through a built-in metalanguage (indexing language) in order to promote the effective retrieval of information - focused on the databases of specialized scientific areas.

According to Fujita, Rubi and Boccato (2009), while subject cataloging is essentially linked to the construction of library catalogs, indexing is linked to the construction of indexes of bibliographies in bibliographic information services that produce databases, as

[...] in book cataloging, its content is treated as a whole, and subjects are provided on a limited scale (a classification number for shelf arrangement and one or two subject headings for access through the catalog). On the other hand, in the indexing of other materials, the tendency is the detailing, in which there is more generosity in the provision of terms for the access by subject (FUJITA; RUBI; BOCCATO, 2009, p. 31).

Although distinct Information Subject Treatment processes, the quality of both subject cataloging and indexing is related to the ability to reconstruct the subject treated in a document into concepts for later retrieval by the user of the information system. The objective of the professional practice of both the indexer and the subject cataloger is to reveal the subject of a document, that is, to represent its content through meaningful concepts.

As the ones responsible for storing, locating, searching and selecting, on demand, data relevant to a particular subject, online catalogs and university library databases are considered information retrieval systems, in which the quality of thematic representation is directly related to the quality of the thematic analysis, considering its inherent complexity in professional doing (TARTAROTTI, 2014, p. 23).

Thus, the objective of the research was to investigate the librarian performance in information retrieval systems in Health, to observe the cognitive processes of analysis of the thesis-type document in subject cataloging in the online collective catalog DEDALUS (USP) and the indexing process in LILACS database (BIREME). Through theoretical-methodological contributions, it was possible to observe if the Information Subject Treatment librarian performs the subject analysis in the process of subject cataloging in collective catalogs, differently from what the professional does in indexing in databases.

Methodological procedures

As a qualitative methodology for data collection, the Verbal Protocol (VP) technique was used in the Individual Verbal Protocol (IVP) modality to observe the librarian's performance. The technique, whose precursors are Ericsson and Simon (1993), consists of a method of introspective analysis whose objective is the recording of the verbal exteriorization of the thought during the accomplishment of a certain reading activity. The subject's "thinking out loud" is recorded and transcribed literally, producing verbal protocols.

The research universe included three Public University Libraries of São Paulo University (USP): Faculty of Dentistry (FO) Library, Faculty of Public Health (FSP) Library and the Institute of Psychology (IP) Library. The sample of three university libraries was defined after analyzing the information retrieval systems and contact with the participating libraries, verifying if they performed both the cataloging of these subject in the online catalog DEDALUS and indexing in the LILACS database, so that the analysis could provide a comparison of the procedures and difficulties surrounding these activities.

In the case of the Faculty of Public Health (FSP) Library, the same professional carried out both the subject cataloging in the online collective catalog DEDALUS and the indexing in the LILACS database, and therefore two protocols were performed with the same librarian.

Regarding the material infrastructure for the implementation of the IVP, the libraries met the necessary requirements. To carry out the protocol, the working environment of the cataloging and indexing librarian was used, so that data collection was able to reflect the reality of cataloging and indexing activities without interference.

Data collection was performed only once in each university library, which resulted six verbal protocols, with five participants. The IVP technique made it possible to compare the professional practice of subject cataloging in an online collective catalog and the indexing process in a specialized scientific database.

Following, the previous procedures are summarized, during and after the application of the IVP:

Procedures prior to the application of IVP

- A) Definition of the research universe;
- B) Selection of participants;
- C) Definition of tasks;
- D) Selection of the base text;
- E) Informal conversation with the participants;
- F) Familiarization of the subjects with the IVP technique.

Procedures during the application of the IVP for subject cataloging/indexing

- A) Recording the participants' "Thinking out loud".

Procedures after the IVP

- A) Literal transcription of the recordings of the participants' speeches;
- B) Detailed reading of the data in search of significant and recurrent phenomena for the elaboration of categories of analysis;
- C) Elaboration of the categories of analysis;
- D) Return to data to remove excerpts from the discussion that exemplify each category of analysis.

Results

The analysis of the results was based on two categories of analysis: Subject analysis/Identification of Concepts and Representation/Translation of Concepts. In relation to the category Subject Analysis/Identification of concepts, it was possible to observe that all the professionals, whether subject catalogers or indexers, carried out the analysis of the document content, however, in different levels of depth. They based on certain parts of the document (textual structure), in search of the identification and selection of concepts and, sometimes, in search of understanding the document's content.

TABLE 1: Parts of text used by professionals during subject analysis/identification of thesis concepts.

CAT - FO	IND - FO	CAT - FSP	IND - FSP	CAT - IP	IND - IP
Title	Title	Title			Title

Abstract	Abstract	Abstract	Abstract	Abstract	Abstract
	Abstract keywords				
Index CardKeywo rds				Index CardKeywo rds	
Introduction					
Figures					
			Summary		

Source: by the authors

In search of meaningful concepts that could represent the content of the thesis-type document, it was verified that the subject cataloger from FO used the following parts of the text: title, abstract, keywords of the index card, introduction and figures, emphasizing, during the IVP, the importance of the technical reading, while the indexer chose the title, the abstract and the keywords of the abstract. In addition, the indexer explained that in her library there are trainees in the area of Dentistry to help professionals as to the terms used.

In FSP, the subject cataloger used the title and the abstract. The same professional, who also carried out the indexing process, used the abstract and the summary of the thesis. Since the two activities were carried out in sequence, it is possible that during the indexing process, the professional considered it unnecessary to re-analyze the title of the thesis. In the IP, the subject cataloger quickly used the abstract to confirm the keywords already assigned in the thesis catalog, while the indexer used the title and abstract to identify the relevant concepts that represented the content of the document.

Regarding the professional guidelines for subject analysis (ISO 5963 for Indexing and Dewey Decimal Classification for Classification), Mai (1997, p. 60) warns that nothing is declared about the examination of the thematic content of documents, indicating only the potential sources to locate the subject. Moreover, none of these common orientations clarifies how the indexer should determine subject content based on the analysis of users' needs and socio-cultural contexts.

According to Standard 12.676 (ABNT, 1992), a translation of Standard ISO 5963, the parts of the text that should have more relevance are: title and subtitle; abstract if any; summary; introduction; illustrations, diagrams, tables and their explanatory titles; words or groups of highlighted words (underlined, printed in different type, etc.); and bibliographic references. However, it is important to highlight the failure

of this standard as it does not identify the types of concepts that can be found in each indicated textual part. According to Fujita (2003), in addition to the indications that the standard makes, it is possible that the professional also uses the first sentences of the chapters to identify the concepts representative of the documents.

All the professionals in the researched libraries used the abstract as the main part of the text to identify and select the concepts representative of the thesis-type document. In this way, it is understood that the abstract is an essential part of this process, preceding even the title at importance level. A well-written abstract should highlight the following points in a concise way: contextualization of the research, rationale, problem, general objective and specific objective(s), results and conclusions. Other parts of the textual structure of theses, such as title, keywords attributed by the author, summary, introduction, illustrations (figures, graphs and tables) as well as the conclusion are relevant elements for the identification and selection of concepts.

Regarding the Representation/translation of concept category, the FO cataloguer explained that the *Sistema de Sugestões do Vocabulário Controlado do SIBi/USP* (SIBIX) does not reach the level of specificity addressed in the thesis. Thus, terms that could be attributed to the document could not be used due to the restriction of the documentary language used in the stage of representation and translation of concepts. However, the same situation occurred with the indexer of the same institution, because when trying to translate into the indexing language, terms identified and selected during the subject analysis to DeCS, the terms were not found.

The FSP cataloger denoted that the same number of terms identified and selected would not be used in the translation stage. The IP cataloger, when analyzing the catalog of the thesis, explained that, through her professional experience, it was possible to verify if the terms contained therein are part of the controlled vocabulary, i.e. SIBIX.

The IP indexer, while identifying and selecting the meaningful concepts of the content of the thesis, also carried out the search of the descriptors in the specific indexing language of the area of Psychology. The indexing language also helped the professional in understanding the documentary content.

Conclusions

The documentary reading is considered the most important phase of the document analysis process, as the success in this initial phase will guarantee the quality of the rest of the process. The difficulty presented by the indexer during the documentary reading in relation to the "identification and selection of concepts representative of the subject of the document" is known to professionals (SILVA; FUJITA, 2004).

[...] the representation by concepts in subject analysis for identification and selection of concepts is different from the representation made in the translation because, first, it is carried out during the analysis of content of the document, that is, with the document in full and secondly, because it uses concepts with which the terms are represented and thus the intrinsic thematics of the document are identified. In the translation, on the other hand, representation occurs with the terms extracted from the analyzed document, therefore, outside the documentary context in which the author developed the meaningful content and uses a documentary language that represents the vocabulary of areas of specialty aiming to match the terms identified with the terms searched by the user (FUJITA, 2013, p. 50, our translation).

Regarding the process, it is understood that during the subject/indexing, the steps do not necessarily have to be performed sequentially, since the professional indexer already familiar with the process can perform them simultaneously (FUJITA; RUBI; BOCCATO, 2009, p.24). In the same way, the more familiar with the specialized scientific area in which one operates, the more these steps are superimposed during the subject cataloging process by the professional in the scope of university libraries.

According to Mai (2000, page 280), the action taken by indexers with different levels of experience may not be exactly the same, but the steps and elements of the subject indexing process are considered fundamental to any indexing process. In this investigation, it was also verified that the professional carried out the subject cataloging and the indexing in a different way, since in the databases the indexing process was carried out in a higher level of specificity than the subject cataloging in the online collective catalog.

As for the subject cataloger/indexer, both the online collective catalog and the databases of specialized scientific areas allow the visibility of the informational content of university libraries. However, the determination of the subject(s) in these information retrieval systems of university libraries involves subjectivity and professional good

sense, considering that both the subject cataloger and the indexer are usually not specialists in the specialized scientific area in which they work.

In relation to the context of the researched libraries, the need for an indexing policy formalized in an indexing manual that guides both the subject cataloging in the online collective catalog and the indexing in specialized scientific databases, whose records are inserted by the professionals working with them is observed. Well-established cataloging and indexing guidelines and procedures followed by professionals can contribute to improving the quality of retrieved information.

Finally, when we look at the professional that performs thematic information processing in university libraries, we contribute to alleviate the incoherence or omissions in the processes of subject cataloging and indexing, aiming at a greater consistency in the creation of informational products.

References

The references was made following the ABNT rules.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. *NBR 12676: Métodos para análise de documentos - determinação de seus assuntos e seleção de termos de indexação*. Rio de Janeiro, 1992. 4 p.

BARITÉ, M. Referenciales teóricos vigentes en el área de tratamiento temático de la información y su expresión metodológica. Porto Alegre: *ABEBD*, 1998. 7 p. [Relatório técnico do II Encontro de Dirigentes dos cursos superiores de Biblioteconomia dos países do Mercosul, Buenos Aires, nov. 1997].

ERICSSON, K. A.; SIMON, H. A. *Protocol analysis: verbal reports as data*. Rev. ed. Cambridge: MIT Press, 1993. 496 p.

FUJITA, M. S. L. A identificação de conceitos no processo de análise de assunto para indexação. *Revista Digital de Biblioteconomia e Ciência da Informação*. Campinas, v.1, n.1, p.60-90, jul./dez. 2003.

FUJITA, M. S. L.; RUBI, M. P.; BOCCATO, V. R. C. As diferentes perspectivas teóricas e metodológicas sobre indexação e catalogação de assuntos. In: FUJITA, M. S. L. (Org.). *A indexação de livros: a percepção de catalogadores e usuários de bibliotecas universitárias*. São Paulo: Cultura Acadêmica, 2009. p. 19-42.

GUIMARAES, J. A. C. Abordagens teóricas de tratamento temático da informação (TTI): catalogação de assunto, indexação e análise documental. *Ibersid*, Zaragoza, p. 105-117, 2009.

ISO. *Documentation: methods for examining documents, determining their subjects, and selecting indexing terms*. Suíça: ISO, 1985. 5 p.

MAI, J-E. The concept of subject: on problems in indexing. In: *Knowledge Organization for Information Retrieval*. Proceedings of the 6th International Study Conference on Classification Research, v. 6, p. 60-67, 1997.

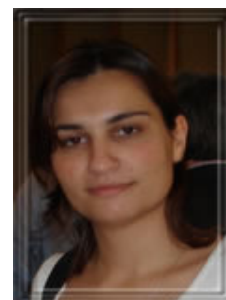
MAI, J-E. Deconstructing the indexing process. *Advances in Librarianship*, v. 23, p. 269-298, 2000.

TARTAROTTI, R. C. D. E. *Atuação bibliotecária no tratamento temático da informação em unidades informacionais: um estudo comparativo qualitativo-quantitativo*. 2014. 277 f. Dissertação (Mestrado em Ciência, Tecnologia e Sociedade) – Universidade Federal de São Carlos - UFSCar, São Carlos, 2014.

The contribution of archival identification to knowledge organization in personal archives



Gabrieli Aparecida da Fonseca
São Paulo State University
E-mail: gabrieli.arq@gmail.com



Sonia Maria Troitiño Rodriguez
São Paulo State University
Email: smtr@marilia.unesp.br

Introduction

Archival identification, as a preliminary step to classification, is an archival procedure of paramount importance for the organization of documentary funds and, consequently, for the knowledge organization emanating from them. However, despite this study discuss documentary genesis, in search of the informational ballast provided by the provenance principle, it is also related to fields of similar studies, such as documentary typology. In this way, it is a study that seeks to articulate information from the understanding of the context of production and accumulation of documents allied to typological analysis, in search of the meanings assumed by the recorded information and ratified by the organicity in its environment.

Considering that in order to explain the origin of the documents it is necessary to go beyond the study of the administrative history of the institution and also to approach the documentary production system through the study of the series derived from it, archival identification can contribute a lot to the effective application of the Principle of Provenance.

According to Ana Célia Rodrigues, archival identification

[...] is a preliminary and necessary task for the functions of classification, evaluation, description and planning of documentary production. [...] In this sense, it is a work of research and criticism on the documentary genesis (RODRIGUES, 2011, p. 118-119).

However, it is still an overlooked area of study, therefore most research developed on the subject generally starts with case studies on public funds. As a consequence, the diverse research on archival identification not always present consensual positions in the field of Information Science – and the studies mainly developed by Spanish researchers, such as Concepción Mendo Carmona or María Luísa Conde Villaverde, among others.

On the other hand, in Archival Science there are several investigative works whose subject falls on the issues of personal archives. However, this research proposes an unusual approach, with the theme of personal archives, aiming to understand how archival identification can contribute to the organization of information from private collections, including when discussing the origin and forms of accumulation that make such archives and their respective documents be personal. In this way, we incorporate to the heart of the discussion the concept Total Archives, defended by the author Terry Cook (1998), who discusses the benefits of informational integration between public and personal archives.

Given these assumptions, we seek to discuss how archival identification can contribute to highlight the boundaries that separate private archives into personal archives and institutional archives. Thus, the case study on the procedures of information organization developed at Fernando Henrique Cardoso Institute - iFHC complements our research when analyzing the documentary treatment of Fernando Henrique Cardoso Fund, in search of the foundation provided by the praxis in archives allied to the theoretical bases of our field of study.

Fernando Henrique Cardoso Institute was created in 2004 by the former President Fernando Henrique Cardoso, in order to gather all his private documentation, as well as his wife's, sociologist Ruth Cardoso. The documents in the custody of the institute are extremely important because, although not of personal origin, they represent part of the history and memory of the country. Among the various documentary types found in the collection are letters, notes, books, texts, photographs, travel notes, technical notes, reports, diaries, certificates, etc.

The election of this collection for the analysis is due to the methodological proposal of the archival organization chosen by the institute. Because the archival identification is a recent field of study and little explored, few institutions incorporate it conceptually in their methodology of work. However, iFHC can be considered a mark

due to the organization work done with the application of archival identification. The pioneerism of the method employed was reported in the book *Tempo e Circunstância: a abordagem conceitual dos arquivos pessoais*, by the authors Ana Maria de Almeida Camargo and Silvana Goulart (2007). The book brings a reflection about personal archives, from the exposition of the organization work realized by iFHC, especially with regard to the archival identification and description.

The concepts and delimitations that differentiate public archives from private, whether institutional or personal, are highly controversial. These divergences are intensified when it comes to politicians' collections, since these are usually constituted both by private documents and by documents created during the public office work. In this sense, archival identification can collaborate to evidence the private nature or not of the analyzed document.

Total archives and archival identification

Total Archives, a Canadian theory that aims to integrate private archives into public archives in order to complement them, arises from the understanding that different contexts of documentary production are equally representative of the society from which they occur. According to Terry Cook, one of the most important researchers on Total Archives, there are two similarities between Personal Archives and Public Archives: first both are produced as a result of records of activities, resulting from human social life. Second, both in public and personal archives the methods and techniques used by archivists to describe, preserve and give access to documents are very similar (COOK, 1998, p. 131)

In this perspective, there is no distancing, but rather approximation and intercommunication between official records and those produced parallel to them, in private. In this way, it is understood that personal archives complement public institutions and vice versa.

According to what Total Archives suggests, personal archives represent the filling of informational gaps in public and institutional archives, because it reveals the documentation that has been hidden because it refers in the foreground to the particularities of the individual, but when articulated to other records, they recover subliminal relationships and give new meaning to the recorded facts in a timely manner. This is due to the fact that the individual, as a societal being, interacts

simultaneously throughout his life, both in the public and private spheres, the documents resulting from this interaction being representative of his own performance.

When reflecting on the role of archival identification in the knowledge organization process, in line with the perspective provided by Total Archives, we could verify that archival identification is crucial for the accomplishment of a satisfactory organization of personal archives, similarly to institutional archives: firstly because of its ability to evidence the nature and origin of documents, according to provenance; secondly, for assisting in the construction of the classification plan by clarifying the organic relations between the different hierarchical levels; and finally, for contributing to the establishment of typological documentary series, through the recognition of species and functions present in the records of human actions in the private sphere.

Finally, we share the position of Fonseca (2015) when she affirms that identification is a means of enhancing documentary treatment, since it anticipates issues necessary to the work performed after it.

Conclusions

Reflecting on the informational values contained in the personal documents of the intellectual and former President Fernando Henrique Cardoso, it was possible to verify that Archival identification can be used in the organization of personal archives in order to reveal the specificities of these collections, through the definition of documentary types, provenance and accumulation of documents. In the case of the analyzed organization, iFHC, the application of archival identification in documentary treatment demonstrated positive aspects, especially for the contribution provided to other archival functions, such as classification and description, both operationally and semantically.

Even today, the concepts and delimitations that differentiate public archives from private ones, whether institutional or personal, are highly controversial, since the characteristics peculiar to each of them are confused. Thus, research on personal and institutional archives are necessary in order to punctuate how archival identification can contribute to a better organization and valorization of this type of documentation, through the study of the genesis and documentary typology.

Ana Maria de Almeida Camargo (2009) shows the necessary care when working with personal archives. According to the author,

If the use of universal labels for the characterization of these files prepares dangerous pitfalls for the professionals who use them - putting them on the same plane, species, forms, genres, subjects and formats - has two other perverse effects: it compromises its organicity and signals the waiving of the evidentiary nature that their original functionality gives them (CAMARGO, 2009, p. 31).

The importance of identification is thus exposed, since it provides an understanding of documents beyond their forms, genres, subjects and formats, as cited, in order to meet their production and accumulation. It thus fulfills its role as an instrument for knowledge organization.

However, the identification work can be much more complex in personal archives. Although the similarities between private and personal archives are large, the latter is characterized by the fact that it presents a large number of non-diplomatic documentary types of the intimate facet, such as letters, postcards, personal journals, among others, as opposed to institutional ones .

The specificities that make the organization of personal archives more complex should not hinder or prevent them from being treated as archives that are - according to the precepts of Archival Science.

Thus, archival identification in personal archives, although complex due to lack of consensus in the area, becomes useful and necessary, as through archival identification it is possible to evidence the provenance and organicity, facilitating the later stages of information organization.

References

The references was made following the ABNT rules.

BRASIL, Congresso Nacional. *Lei nº 8.394, de 30 de dezembro de 1991*. Diário Oficial da União, de 6 de janeiro de 1992. Disponível em:
<<http://www.conarq.arquivonacional.gov.br/cgi/cgilua.exe/sys/start.htm?from%5Finfo%5Findex=11&infoid=97&sid=52>> Acesso em : 16 de dez. de 2013.

CAMARGO, Ana Maria de Almeida. Arquivos pessoais são arquivos. *Revista do Arquivo público Mineiro*, p. 27-39. Disponível em:
<http://www.siaapm.cultura.mg.gov.br/acervo/rapm_pdf/2009-2-A02.pdf> Acesso em: 18 de jul de 2013.

CAMARGO, Ana Maria de Almeida; GOULART, Silvana. *Tempo e circunstância: a abordagem contextual dos arquivos pessoais-procedimentos metodológicos adotados na organização dos documentos de Fernando Henrique Cardoso*. São Paulo: Instituto Fernando Henrique Cardoso, 2007.

CARMONA MENDO, Concepción. Consideraciones sobre el concepto de documento privado. *SIGNO. Revista de Historia de la Cultura Escrita*, v. 3, p.11-23, 1996, Universidad de Alcalá de Henares. Disponível em: <http://dspace.uah.es/dspace/bitstream/handle/10017/7489/consideraciones_mendo_SIGNO_1996.pdf?sequence=1> Acesso em: 12 de jun de 2014.

CARMONA MENDO, Concepción. Consideraciones sobre el método en archivística. In: *Documenta & Instrumenta*. Madrid, Universidad Complutense de Madrid, v.1, p. 35-46, 2004. Disponível em: <<http://revistas.ucm.es/index.php/DOCU/article/view/DOCU0404110035A/19190>> Acesso em: 11 de jun de 2014.

CONDE VILLAVERDE, Maria Luísa. La investigación em los archivos. Evolución de su contexto y contenido. *ARBOR*, v.CLXXXII, n.717, p. 33-37, 2006. Disponível em: <<http://arbor.revistas.csic.es/index.php/arbor/article/view/5/5>> Acesso em: 01 de dez. de 2013.

COOK, Terry. Arquivos Pessoais e Arquivos Institucionais: para um Entendimento Arquivístico Comum da Formação da Memória em um Mundo Pós- Moderno. *Revista Estudos Históricas*, v. 11, n. 21, p. 129-150, 1998 Disponível em:<<http://bibliotecadigital.fgv.br/ojs/index.php/reh/article/view/2062/1201>>Acesso em: 11 de ago de 2013.

FONSECA, Gabrieli Aparecida da. *A contribuição da identificação documental em arquivos pessoais: um estudo de caso no Instituto Fernando Henrique Cardoso*. Marília, 2015.

RODRIGUES, Ana Célia. Identificação: uma nova função arquivística. *Revista EDICIC*, v,1, n.4, p. 109- 129, 2011. Disponível em: <<http://www.edicic.org/revista/>> Acesso em: 08 de abr de 2013.

Document analysis and the generative trajectory of meaning in the representation of the archival document



Gilberto Gomes Cândido
São Paulo State University
ggcandido@marilia.unesp.br



João Batista Ernesto de Moraes
São Paulo State University
jota@marilia.unesp.br



Deise Sabbag
University of São Paulo
deisesabbag@usp.br

Introduction

Discussions have been raised on the representation of the informational content of archival documents, since neither the General International Standard Archival Description (ISAD (G)) nor the International Standard Archival Authority Record for Corporate Bodies, Persons and Families (ISAAR (CPF)), compromise with the entry or exit of the information in the Archive environment. Thus, the responsibility is taken by each institution for the elaboration of its research instruments, as well as for the identification of methods that may help in this endeavor.

In order to contribute to this issue, two methodologies are presented as auxiliary tools in the representation of the archival document and in the identification of its

production context: a) the Generative Trajectory of Meaning assisting in the analysis of the textual structure of the document; b) Document Analysis (DA) through the analytical stage, presenting the most relevant points for reading the document.

In the area of Information Science, these methodological proposals have been successfully used to assist in the thematic identification of fiction texts and scientific articles, by several authors, such as: Moraes (2011; 2012); Moraes, Guimarães and Guarido (2007); Moraes, Damazo and Lara (2008); Moraes and Guimarães (2008); Sabbag (2012; 2013); Guarido and Moraes (2009); Moraes and Alves (2009); Garcia-Marco, Moraes, Garcia-Marco and Guimarães (2010); Moraes and Candido (2013); Candido, Lima and Moraes (2013) and Candido (2014).

Thus, the research aims at applying the methods for identifying the archival document contents. The corpus of analysis is the archival document belonging to the police station of Ocaçu/SP, district dismembered from Marília/SP in 1959, under the responsibility of Centro de Documentação Histórica e Universitária de Marília – CEDHUM.

Therefore, we seek for contributions in the methodologies to improve the representation of the archival document and its organization, dissemination, access and use of information.

2 Archival Document

According to Jenkinson (1922), to be considered and classified as belonging to the class of archive, the document should have been elaborated in an administration (public or private) as well as transited within it (1922, p.11).

Muller, Feith and Fruin (1973, p.13) extended the definition of archival document by considering, any writings, drawings and printed material, received or produced officially by a particular administrative entity or by one of its officials, as such documents were intended to remain in the custody of that entity or official as evidence of activities.

Schellenberg (2012, p. 41) shows that archival documents are proof of their functions, policy, decisions, methods, operations or other activities, or as a consequence of the informative value of the data contained therein.

Sharing Schellenberg's (2012) understanding, Rousseau and Couture (1998, p. 137) argue that the archival document is a set consisting of a medium and the information it contains, usable for purposes of consultation or as proof.

Thus, the value of evidence given to the archival document is to record routine activities of a public or private administration or social individuals, in the same way as to provide the identification of the trajectories performed for its elaboration, according to the authors, various media to record the information.

However, according to Jenkinson (1922), there is also the inference of two qualities that assure the archival document the value of proof: impartiality concerning the truth of administrative action; and authenticity about the maintenance, integrity and preservation of the file structure.

Complementing Jenkinson (1922), Duranti (1994) exposes three more qualities being considered peculiar to the archival document and that also allow to maintain the value of proof: naturalness, interrelation/organicity and uniqueness. The naturalness of the archival document comes from its uninterrupted elaboration as a reflection of the functions performed by the institutions, in order to prove their activities; the interrelation/organicity occurs in relation each document has to the functions that generated it; " uniqueness, which derives to each archival document by the fact of its having a unique place in the structure of the group in which it belongs and in the documentary universe" (DURANTI, 1994, p. 03).

The management activities of an archive must ensure that these five intrinsic qualities are assigned to the document, guaranteeing the attributes of specificity, so that they can aid in the documentary contextualization, and in the maintenance of its recurrent probative value.

Thus, for the archival document analysis, document analysis was used in its analytical stage as means to carry out the reading in order to highlight the passages of its content.

3 Document analysis

Document Analysis (DA), in the scope of Information Science, according to the authors Gardin et al (1981); Garcia Gutierrez (1984); Cunha (1987); Pinto Molina (1989); Coyaud (1966); Chaumier (1971; 1974); Clauso Garcia (1933); Ruiz Perez (1992); Guimarães (2003); Pinto, Gavez (1999) and Pinto Molina (1991), is a set of procedures that aim at the identification of the contents of documents, and the subsequent synthesis, for the purpose of representation through a secondary document (products: catalog, controlled vocabulary , etc.) considered an instrument that provides access, consultation and information dissemination.

It is noteworthy that Document Analysis is divided into two processes: Documentary Analysis of Form (ADF), a descriptive process of the elements/characters extrinsic to the document; and Document Analysis of Content (ADC), a process of identifying the elements/characters intrinsic to the document (PINTO MOLINA, 1992; RUIZ PÉREZ, 1992; FOX, 2005; NASCIMENTO, L.M.B 2009).

For this study, only the procedures of Document Analysis of Content (ADC) will be used, which is divided into two stages: 1) analytical and 2) synthetic. The first one corresponds to the processes of identification of the documentary typology, technical reading and identification of concepts; the second one, the selection of concepts, documentary condensation and documentary representation.

Among the steps presented, we mainly focus on the analytical stage of the ADC process, more specifically on technical reading, in which the analyst explores the textual structure of the document, identifying the most relevant parts of the content.

This reading procedure comprises a set of cognitive and metacognitive strategies. These strategies are related to the objectives that will be reflected in the course of the documentary reading, whose aspects are addressed by Pinto Molina (1993, pp. 162-163), when he highlights that strategies are developed during the reading period and not depend only on the reader and the text, but also on the documentary objectives. These operations are mental actions that work in organized ways, and can be evidenced in two categories. The first one, cognitive, which comprises interpretation as automatic and unconscious behavior; the second one, the metacognitive, which presupposes unattached activities, in which the reader is aware of how he is reading (PINTO MOLINA, 1993, pp. 162-163).

In this context, reading resources contain conceptual linguistic and cognitive characteristics, being used in information processing through the bottom-up and top-down approaches, which are considered as coming from cognitive psychology.

Pinto Molina (1993, p. 161) shows that the bottom-up approach is the linear type of reading, going from the parts to the textual whole, which is based exclusively on visual stimulation, while the top-down approach is the method of deductive reading, in which one walks in an inverse way, altogether for the parties, taking advantage of the 'prior knowledge' or 'schemes'.

Thus, based on the author, it is understood that reading methods are dynamic and interpretive; their understanding is established by the actions of meaning, in such a way to adapt the semantic analysis and content analysis.

4 Generative trajectory of meaning

The Generative Trajectory of Meaning results from the studies of Greimas (1917-1992), who sought, in textual structures, means to explain the development and meaning of the written text.

This trajectory is understood as a theoretical expression of semiotics, which seeks, in its components, means of articulation that promote its interaction, so as to present a generation perspective, that is, assuming that every semiotic object can be defined in its own production way, therefore the elements that permeate this process are correlated with each other forming a "trajectory" that goes from the simplest one to the most complex one, from the most abstract one to the most concrete one"(GREIMAS, 2008: 232) .

According to Fiorin (1999, p.17), the generative trajectory of meaning has three levels of structures: the fundamental ones, the narrative ones and the discursive ones. The fundamental or deep structure is understood as the semantic category; the narrative, as canonical structure; and the discursive, as concretization of themes and figures. The methods of analysis of the three levels are different from each other, but a correlation occurs between them, so that one has the full understanding of the text, starting from the simple to a more complex analysis.

Examples of the levels are the following:

- Fundamental or Deep structure: it is established by a relation of setbacks, according to Fiorin (1999, p.10) it comprises the semantic category or categories that order, generally, the different content of a text. A semantic category is an opposition such as a versus b. In the archival document, this can be observed by terms that present this level of contrariety, for example: Deferred x Indeferred, Ratify x Rectify, etc. , with this analysis, it is possible to identify simple and abstract features of the text, as well as its semantic opposition;

- Narrative structure: occurs through the interlocution between the subject and the object connected to be recognized narrating a fact/a story. It is noteworthy that the archival document is understood as descriptive, that is, objective; when describing a function and activity. At this level we have the canonical sequence of the text: **Manipulation** (willing or duty), **Competence** (knowing or being able to do), **Performance** (transformation caused by a change from a state to another), and **Sanction** (distribution of prizes and punishments);

- Discursive structure: it indicates the concrete elements of the text through the terms that refer to themes and figures, the theme contains predictive and interpretative characteristics of the natural world, and the figure simulates the natural world through description.

However, for the thematic and figurative trajectory to be structured, it is indispensable to use the isotope which, according to Greimas and Courtés (2008, p. 278), it is a reading screen that makes the surface of the text homogeneous, since it allows avoiding ambiguities. According to the authors, it is through isotopes that we identify the semantic sequences that converts the text read into a set.

Complementing this prospect, Moraes and Guimarães (2006, p.07), demonstrate that in the stage of "discursive structures, the results obtained in this phase should refer to the initial stage, that is, to the fundamental structures", that is, the logical path of the construction of a coherent text.

It should be noted that, in the Generative Trajectory of Meaning method, we used only the narrative and discursive structures, together with AD: the first one by extracting the canonical sequence of the text; the second one by allowing the analysis of the themes and figures of each sequence.

5 Application of the generative trajectory of meaning in the textual structure of the archival document

The application of the generative trajectory of meaning was on the archival document, chosen at random from Centro de Documentação Histórica e Universitária de Marília – CEDHUM, the document belonging to the police station of the district of Ocaçu/SP, district dismembered from Marília/SP in 1959.

Documentary reading was carried out in the document, which enabled the identification of the narrative level and the extraction of the textual elements present in the canonical sequence: manipulation, competence, performance and sanction.

After the identification of the canonical sequence, the discursive level was verified, compiling the themes and figures that pointed out the concrete elements of the document.

The experiment follows:

Chart 1 - Manipulation - Canonical Structure Element

Source: by the authors

MANIPULATION
In obedience to the record 1368 by Mr. Deputy Delegate of this Regional Police Station, I performed a small inquiry into the issue of the heifer that had been improperly appropriated by José Oliveira, which belongs to Mr. Pedro dos Santos.

Chart 2 - Figurative Trajectory and Thematic Trajectory

Themes	Figures
	Mr. Deputy
Inquiry	Regional Police Station
	Heifer
Appropriated	José Oliveira
Property	Pedro dos Santos

Source: by the authors

Chart 3 - Competence - Canonical Structure Element

COMPETENCE
In the course of the interrogatory, I could not ascertain anything, nor could I get an understanding between them. As seen from José Santini's testimony, page 5, and his son-in-law João de Oliveira's statement, page 4, there is a certain contradiction, since João de Oliveira, owner of the farm, confesses that he had a heifer owned by Pedro Bernardo when he removed the cattle, while Santini, the keeper of his son-in-law's farm, denies keeping any cattle owned by Bernardo. The testimony of the witnesses are all in favor of Pedro Bernardo with the exception of José Ribeiro de Paulo, who sold him the cattle in question.

THEMES	FIGURES
Interrogatories	
Statements	José Santini
Declaration	João de Oliveira
	Farm
Confess	Heifer
Removed	Cattle
	Pedro Bernardo

Source: by the authors

Chart 4 - Figurative Trajectory and Thematic Trajectory

PERFORMANCE
In this condition, I send the statements of the witnesses to this Regional, so that the case is duly investigated there.

Source: by the authors

Chart 5 - Performance - Canonical Structure Element

THEMES	FIGURES
Statements	Witness
Ascertained	Regional

Source: by the authors

Chart 7 - Sanction - Canonical Structure Element

SANCTION
According to what was verified from Pedro Bernardo's and other witnesses's statements, about 2 months ago, a cow belonging to Pedro Bernardo crossed to João de Oliveira's, and there it was ironed with the farm's mark, and was given back to Pedro Bernardo.

Source: by the authors

Chart 8 - Figurative Trajectory and Thematic Trajectory

TEMAS	FIGURAS
Statements	Pedro Bernardo
Several	Witnesses
Cause	Cow
Ironed	João de Oliveira
Marked	Farm

Source: by the authors

In the procedures of classifying themes and figures, the structure of the thematic trajectory and the figurative trajectory were obtained as a result, for the establishment of themes.

It is observed that the figures and themes, when analyzed individually, are not relevant, but at the moment the conceptual analysis occurs, we can verify the functioning and the construction of the text with the support of the question of isotope, as a procedure of construction of the meaning on the syntagmatic dimension, which controls the meanings of the words before the context.

Thus, the analysis of the text presented the term "Inquiry" as a semic trace, which tends to reproduce through isotope. Therefore, the Generative Trajectory of Meaning helps us to understand the textual content of the Archive Document, since this technique breaks the text into sections that, with interpretation, provide the identification of the narrative circuit, which relates it to the function that elaborated it for a given activity.

However, we highlight that, such applied method works with the textual content or its residue regardless of its medium. In this way, we believe that this method would not be usual, in documentary species, such as: notes of commitment, delivery reports, balance sheets etc.

6 Conclusions

The Archival Document is descriptive and objective, however, as a result of its representation, it can also be perceived as having a narrative content, in which a degree of subjectivity exists because of the interpretation intention, and in this case the archivist becomes a storyteller.

Duff and Harris (2002, p.276) point out that, at the moment of representing the archival document, certain elements/characters are selected, which can then be used to represent it. However, " In this process, there is analysis, listing, reproduction, and so on, but its primary medium is narrative.", that is, the act of telling a story (DUFF; HARRIS, 2002, 276).

Therefore, the Generative Trajectory of Meaning is shown, in a punctual way, for representing the information of the archival document, helping in the identification of the function and activity, in order to provide a better information representation, dissemination, access and use.

However, it was also pointed out that the Generative Trajectory of Meaning, in certain documentary species, such as: notes of commitment, delivery reports, balance sheets, etc., would not be usual.

References

The references was made following the ABNT rules

DUFF, W. M.; HARRIS, V. Stories and Names: Archival Description as Narrating Records and Constructing Meanings. *Archival Science*, v. 2, n. 3-4, p. 263-285, 2002. Disponível em: < <http://yalearchivalreadinggroup.pbworks.com/f/Duff&Harris.pdf>>. Acesso em: 15 de Jul. 2014.

DURANTI, L. The concept of appraisal and archival theory. *The American archivist*. Chicago: *The Society of American Archivists*, v.57, n°2, 1994a, p.328- 344. Disponível em: < https://courses.ischool.utexas.edu/Winget_Megan/2007/Summer/Becker/Duranti_Concept_of_Appraisal.pdf> Acesso em: 16 Nov. 2013.

FIORIN, J. L. *Sendas e veredas da semiótica narrativa e discursiva*. DELTA. fev./jul. 1999, vol.15, no.1. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-44501999000100009&lng=pt&nrm=iso. >. Acesso em: 16 Jan. 2014.

FLOCH, J. M. *Documento de Estudo do Centro de Pesquisa Socios semióticos*. Periódicos I. São Paulo: Centro de Pesquisas Socios semióticos, 2001.

FONSECA, M. O. *Arquivologia e Ciência da informação*. Rio de Janeiro: FGV, 2005. p. 121.

FOX, V. *Análisis documental de contenido: principios y prácticas*. Buenos Aires: Alfagrama, 2005.

FUJITA, M. S. L. A leitura documentária na perspectiva de suas variáveis: leitor-texto-contexto. *DatagramaZero*: revista de Ciência da Informação, Rio de Janeiro, v. 5, n. 4, ago. 2004. Disponível em: <http://www.dgz.org.br/ago04/Art_01.htm>. Acesso em: 12 Mai. 2014.

GREIMAS, A. J.; COURTÉS, J. *Dicionário de Semiótica*. São Paulo: Cultrix, 2008.

GUIMARÃES, J. A. C. *Análise documentária em jurisprudência: elementos para uma metodologia de indexação de acórdãos trabalhistas brasileiros*. 1994. 250 f. Tese (Doutorado em Ciências da Comunicação) – Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo, 1994.

JENKINSON, H. *A manual of archive administration: including the problems of war archives and archive making*. London, Edinburgh, New York, Toronto, Melbourne and Bombay: Oxford at the Clarendon Press, 1922. 281 p. Disponível em: <<https://ia700302.us.archive.org/0/items/manualofarchivea00jenkuoft/manualofarchivea00jenkuoft.pdf>>. Acesso em: 18 Dez. 2013

MORAES, J. B. E.; GUIMARÃES, J.A.C. Análisis documental de contenido de textos literarios narrativos: en busca del diálogo entre las concepciones de aboutness/meaning y de recorrido temático/recorrido figurativo. *Scire* (Zaragoza), 2006.

NASCIMENTO, L. M. B. *Análise documental e análise diplomática: perspectivas de interlocução de procedimentos*. 2009. 198 f. 2009. Tese de Doutorado. Tese (Doutorado em Ciência da Informação) - Universidade Estadual Paulista, São Paulo.

PINTO MOLINA, M. *El resumen documental: principios y métodos*. Madrid; Salamanca: Fundación Germán Sánchez Ruipérez; Madrid: Pirámide, 1992.

PINTO MOLINA, M. *Análisis documental: fundamentos y procedimientos*. 2. ed. rev. aum. Madrid; Eudema, 1993.

ROUSSEAU, Jean-Yves; COUTURE, C. *Os fundamentos da disciplina arquivística*. Lisboa: Publicações Dom Quixote, 1998.

RUIZ PÉREZ, R. *El análisis documental: bases terminológicas, conceptualización y estructura operativa*. Granada: Ed. Universidad de Granada, 1992.

SCHELLENBERG, T. R. *Arquivos modernos: princípios e técnicas*. Tradução de Nilza Teixeira Soares. Rio de Janeiro: Editora FGV, 2012, p.320.

Contribution of the Generative Trajectory of Meaning for documentary reading of narrative texts of fiction



Deise Sabbag
University of São Paulo
deisesabbag@usp.br



João Batista Ernesto de Moraes
São Paulo State University
jota@marilia.unesp.br

Introduction

Documentary reading is an important activity within the analytical stage of Document Analysis, enabling the identification of the most relevant parts of the text so that the analyst can perform the identification of concepts. In this sense, we observe that several documentary typologies offer a textual structure that favors the applicability of reading strategies, such as technical-scientific documents that have the pre-textual, textual and post-textual elements.

We also observe that the use of the same reading strategies for the analysis of narrative texts of fiction need to be studied so that this narrative typology is not only represented by its genre or nationality.

In this sense, this theoretical study, from a bibliographic method, used the Generative Trajectory of Meaning to present the underlying regularities of fiction narratives that allow the representation of document contents, aiming at proposing a strategy for documentary reading of narrative texts of fiction.

The paper is organized in three parts that present the definition of narrative texts of fiction to the reader, followed by the theoretical subsidies of the area on documentary reading and, finally, the proposal of documentary reading of narrative texts of fiction having the Generative Trajectory of Meaning as theoretical underpinning.

The narrative text of fiction

Every text has, in its construction, a superficial structure (microstructure), a deep structure (macrostructure) and a social use (superstructure). For a text to be classified as a "narrative text", it is necessary that its microstructure consists basically of elements that indicate actions and changes of state (verbs and indicators of time and place); the macrostructure presents characters that undergo changes of state; and the superstructure is characterized by a sequence of actions.

In this sense, we present a reflection on the establishment of the criteria that define what the fiction narrative text is developed from the literature of the area of Information Science, using definitions provided by Lancaster (2004) and Beghtol (1994), making a juxtaposition as performed by Moraes (2011).

Lancaster (2004) refers to fiction when writing about indexing. The author compares the indexing of fiction texts in counterpart of nonfiction texts, and concludes that these two textual typologies present different issues for indexing. These differences originate from their genesis, since the two typologies are constructed with different goals, fiction for entertainment and nonfiction to convey information.

For Beghtol (1994), the world of documents can be divided into works that are born from the imagination of their creators and works that are born from the rational faculty. For the author, fiction can be classified as a work that arises from the imagination of its creators, being written in narrative prose, "narrative can be taken broadly to include discourses that progress in some sense from one point to another" (BEGHTOL, 1994, P. 7).

In the analysis of the definitions carried out by Moraes (2011), Lancaster (2004) presents a reductive perspective on the literary work, which can be interpreted as

unfamiliarity of the literary manifestations when the author affirms that fiction texts are destined to entertain.

Obviously, these works can serve the fun, provoke emotions, but we argue the dimension of fiction as an artistic manifestation (present and recognized from antiquity) that can not be forgotten, much less neglected in the area of Information Science.

The reductionism of Lancaster's (2004) definition is contrasted by an amplitude in the definition by Beghtol (1994) regarding the fiction text, since this comes from the imagination of its creators. The literary text lies at the threshold of the imagination, the imaginary, and how to establish the limits of 'what is imaginary or not.' For the author, fiction is narrative prose, this definition also contemplates a great number of texts that are not narrative (or only narrative), but does not contemplate texts that can be classified as fiction.

In order to construct a definition of fiction text that is more appropriate to the studies directed to the area, Moraes (2011) takes the statements of Beghtol (1994) on a self-criticism to her definition of fiction as a starting point:

It is useful to accept such a broad operational definition for two reasons. First, no definition of fiction is universally accepted (...). Second, the examination of fiction documents for information storage and retrieval needs to include as many examples of as many types of fiction as possible. (BEGTHOL, 1994, p. 07).

The author states that there is no consensus on what fiction is, pointing out that a definition should be so comprehensive as to encompass all the possibilities of the fiction text (MORAES, 2011, p.20).

In this way, Moraes (2011) presents fiction as an art form, an artistic manifestation constructed with words that surpass the original writings of the writer's imagination, which may contain concrete elements of reality.

In this perspective, the fiction narrative text can be defined as a work of art in which the artist is the writer who creates the terms that represent the context chosen by the semantic content.

To consider the fiction narrative text of as an artistic manifestation, as a work of art, is to throw a totally new look at this textual manifestation, as well as for documentary reading and the entire process of Document Analysis. It is to consider that their interaction, and interlocution, act in a peculiar way in the development of the communicative competence of the analyst.

This communicative competence will provide a specific type of interaction with the reader-analyst: interaction with an artistic creation; interaction with an artistic form; interaction with an artistic expression; providing the knowledge of its textual typology and the establishment of reading strategies.

Documentary reading

Documentary reading composes the process of document analysis. This process has particular characteristics and constraints within Document Analysis (AD).

In the literature, we find other denominations for this activity: professional reading (MOURA, 2004), document technical reading (GUIMARÃES, 1994, DIAS, NAVES, 2007), documentary reading (CINTRA, 1987; LARA, 1993), reading for documentary purposes (LARA, 1993), fast or diagonal reading (CHAUMIER, 1988), careful reading (CAVALCANTI, 1982), professional reading (FUJITA, 2004), mixed of reading and overlooking (LANCASTER, 2004) document examination (NBR 12676), shortcuts (FOSKETT, 1973), reader-librarian (MOURA, 2004) and so on.

Documentary reading presupposes a specific modality of reading in which technical requirements will be used to analyze a variety of documents.

Kato (1985) develops the conception of readability for reading the text, taking into account the interactive view of reading, that is, approaching the act of reading as an interaction between the reader and the text.

Briefly, for Kato (1985), readability, in an interactive view, has, as its function, a good textual formation which takes into account not only the structural question, but also the functionality; the reader's prior knowledge through the operations of recognition and also of inferences of the new through what is known; and the balance between cognitive and metacognitive strategies.

Based on Kato's conception of readability (1985), Cintra (1987) agrees that those factors contribute to the quality of the text, and presents, based on them, which reading strategies would be applied to the professional practice of the librarian.

Cintra (1987), Lara (1993), and Tálamo (1982) use the concept of global reading used by Grice (1975) to highlight the importance of understanding the overall reading process for later reading comprehension.

In this process of global reading, Grice (1982) presents the principle of cooperation, that is, interlocution that occurs between the reader and the text, in which the reader negotiates with the text until reaching a particular interpretation.

According to Cintra (1987), in reading for documentary purposes, the principle of cooperation (author/reader), proposed with Grice (1982) is broken, as the author did not foresee the professional reader in this process.

Thus, as Cintra (1987) points out, considering the prior knowledge and types of strategies that the text requires as legibility factors in an interactive process, is to recognize that more than a set of visual signals are involved in the reading process. In fact, it means to admit that there are "cognitive aspects related to knowledge stored in the reader's memory and specific behaviors during the reading" (CINTRA, 1987, p.29).

Important prior knowledge for document technical reading is the knowledge on typology and textual superstructures existing in any textual construction. Knowing the textual typology of the document, the reader can dispense with reading word for word, once he/she dominates the superstructures, being able to interpret the central ideas of the text. The mastery of textual superstructures allows the reader to create reading strategies or mental actions that lead him/her to achieve his/her goals (CINTRA, 1987, p.31).

According to Cintra (1987), the selection of strategies during reading depends on several factors: the purpose of the reading, the reader's experience, maturity in relation to the reading task, the type of text, the more focused attention in parts of the text, the degree of novelty of the text and the motivation to read. As for the strategies, they can be cognitive (automatic and unconscious behaviors) and metacognitive (as the reader is aware of how they are reading).

For Cintra (1987), the experienced reader would be characterized by the abilities of monitoring, identification of important parts of the text, concentration, text segmentation, corrective actions of detected failures in the process.

Thus, the technical reading of any document, in its most diverse manifestations, allows the construction of substitutes, that is, of documentary representations.

Documentary reading of narrative texts of fiction

Traditionally, in the area of Information Science, the same methods of documentary reading for a wide variety of types of documents are applied, with minor

changes. Some strategies are used for documentary reading, such as: reading the title of the works; subtitles; introduction and conclusion; introductory sentences of paragraphs and chapters; captions of illustrations; graphs; tables; diagrams and their explanations; conclusion; words or groups of words.

These procedures demonstrate efficacy in documents that provide a textual and physical structure suitable for the applicability of such strategies, such as scientific papers, technical reports, textbooks, and so on. But when the same strategies are applied to the documentary reading of a fiction narrative text, the same results are not obtained, that is, the thematic content of the document is not defined.

One solution would be the full reading of narrative fiction texts, but due to issues of time, the professional does not have this possibility in libraries. Thus, what strategic resource should the analyst use for the technical reading of a fiction narrative text?

It is known that the narrative text of fiction as text is composed of a superficial structure (microstructure), which comprises linguistic elements that aid in its composition and by a deep structure (macrostructure), in which the elements that make up the so-called semantics of the text exist, and also by a form defined by social use (superstructure).

The narrative text of fiction has four characteristics that must be present in a text so that it is a narrative: concrete situations, figurativity, relations of posteriority, concomitance and anteriority between the reported episodes and preferential use of the temporal subsystem of the past. It has a content plan designed in the form of a generative path. This gives narrative fiction the concept of content organized in narrative articulations, that is, it has a plan for its construction.

Considering this element, in addition to the notion of the "figure of the spectacle" which is a minimal, condensed structure that can be the basis for several narratives, narrative texts have a constant basic structure; a structure defined by the Generative Trajectory of Meaning.

Through the Generative Trajectory of Meaning necessary elements for the establishment of strategies for the technical reading of narrative texts of fiction are found.

The Generative Trajectory of Meaning possesses levels: fundamental; narrative; discursive and manifestation. The narrative level is composed of the so-called canonical sequence, and this sequence includes the theoretical subsidies

necessary for the establishment of a strategy for documentary reading of narrative texts of fiction.

The canonical sequence has four phases described below (FIORIN, 2011, pp. 29-31):

- **Manipulation**: one subject acts on another to get him to want and/or to do something. The subject is a narrative role and not necessarily a person. There are several types of manipulation, here we describe the four most common ones: temptation, intimidation, seduction and provocation;
- **Competence**: the subject who performs the narrative is endowed with a knowledge and/or power to do;
- **Performance**: phase in which the central transformation of the narrative takes place (change from one state to another). For example: in a fairy tale in which a princess was kidnapped by the dragon, the performance will be the release of the princess. There is, therefore, a subject that operates the transformation and the subject that comes into conjunction, or in disjunction with an object that can be distinct or identical;
- **Sanction**: the last phase when there is the realization that the *performance* has materialized and the recognition of the subject that has made the transformation. The sanction may be cognitive if there is recognition that competence has been realized; or the penalty can be pragmatic, with rewards and punishments.

Thus, one can consider that the stage of **performance** and **sanction** are usually structurally located in the last chapters of the works, following the figure of the spectacle advocated by Vladimir Propp. The figure of the spectacle could be the basis for several narrations, as it presents itself as a minimal structure that can make possible several narratives. Obviously, throughout the narrative situations of performance and sanction also occur, but where the controversial dimension of the narrative is located is, strictly, in the last two chapters. Thus, in narrative fiction texts, conjunctions, or disjunctions, for the main characters usually occur in the final chapters. It also shows the figure of the spectacle, that is, the presence of a minimal structure that can be expressed with the statement "The father gives the son a gift".

Thus, as a strategy for the analysis of narrative texts of fiction, the documentary reading of the last chapters of the works, is proposed as a thematically significant part of the text structure. The application of this strategy will enable the analyst to use the conceptual categories that will allow the construction of subject statements.

Thus, for the examination of narrative texts of fiction we propose the use of the following reading strategies (reading the following parts of the document)

- A) Title of works;
- B) Subtitles;
- C) Abstract;
- D) Summary;
- E) Illustrations, diagrams, tables and explanatory titles (not common in narrative fiction texts); and
- F) Conclusion (composed of the last chapters of the works).

Conclusion

Documentary reading is an important activity within the analytical phase of Document Analysis. In the indexing process, it manifests itself as a stage of fundamental importance for understanding the texts that will be translated and represented.

This type of reading requires certain technical requirements. Among the requirements, we highlight the prior knowledge and types of strategies that the text requires. The previous knowledge that we present in this work, as important for reading narrative texts of fiction, was the knowledge of typologies and textual superstructures.

As a typology, the analyst (indexer) needs to recognize the narrative text of fiction as an artistic manifestation, characterized by the verbal medium, having its artistic purpose composed of words unveiled by the reader's decoding ability. This artistic manifestation provides a new interaction that develops a communicative competence in the analyst. It is also important for the analyst to recognize the textual superstructures of narrative texts of fiction, as the recognition of the superstructures will allow to develop the reading strategies.

These strategies were reached in this work, based on the studies of the Generative Trajectory of Meaning Theory that demonstrated that the narrative text of fiction has an organized content plan with underlying narrative regularities, that is, a textual structure that allows the analyst to identify which the most important parts of the text for documentary reading are and, later, identification of concepts.

References

The references was made following the ABNT rules.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. *NBR 12676: métodos para análise de documentos: determinação de seus assuntos e seleção de termos de indexação*. Rio de Janeiro, 1992.

BEGHTOL, C. *The classification of fiction: the development of a system based on theoretical principles*. Metuchen: Scarecrow Press, 1994.

CAVALCANTI, C. R. Indexação. *Estudos Avançados em Biblioteconomia e Ciência da Informação*, Brasília, DF, v. 1, n. 1, p. 211-233, 1982.

CHAUMIER, J. Indexação: conceito, etapas e instrumentos. *Revista Brasileira de Biblioteconomia e Documentação*, São Paulo, v. 21, n. 1/2, p. 63-79, jan./jun. 1988.

CINTRA, A. M. M. Estratégias de leitura em documentação. In: SMIT, J. W. (Coord.). *Análise documentária: a análise da síntese*. Brasília, DF: MCT; CNPq; IBICT, 1987. p. 28-35.

DIAS, E. W.; NAVES, M. M. L. *Análise de assunto: teoria e prática*. Brasília: Tesaurus, 2007.

FIORIN, J. L. *Elementos de análise do discurso*. 15. ed. São Paulo: Contexto, 2011.

FOSKETT, A. C. *A abordagem temática da informação*. São Paulo: Polígono, 1973.

FUJITA, M. S. L. A leitura documentária na perspectiva de suas variáveis: leitor-texto-contexto. *DatagramaZero: revista de Ciência da Informação*, Rio de Janeiro, v. 5, n. 4, ago. 2004. Disponível em: <http://www.dgz.org.br/ago04/Art_01.htm>. Acesso em: 12 jan. 2012.

GRICE, H. P. Logic and conversation. In: COLE, P., MORGAN, P. *Syntax and semantics 3: speech acts*. New York: Academic Press, p. 41 – 58. Disponível em: <<http://www.ucl.ac.uk/ls/studypacks/Grice-Logic.pdf>>. Acesso em: 2 abril 2013.

GRICE, H. P. Lógica e conversação. In: DASCAL, M. *Fundamentos metodológicos da linguística*. Campinas: 1982. p. 81-103.

GUIMARÃES, J. A. C. Análise documentária no âmbito do tratamento da informação: elementos históricos e conceituais. In: RODRIGUES, G. M.; LOPES, I. L. (Org.). *Organização e representação do conhecimento na perspectiva da Ciência da Informação*. Brasília: Thesaurus, 2003. p. 100-117.

KATO, M. A. Uma visão interativa da legibilidade. *Ilha do Desterro*, Floiranoópolis, n. 13, p. 57-66, 1985.

LANCASTER, F. W. *Indexação e resumos*. Brasília, DF: Briquet de Lemos, 2004.

LARA, M. L. G. de. *A representação documentária: em jogo a significação*. 1993. 132 f. Dissertação (Mestrado em Ciências da Comunicação) – Escola de Comunicação e Artes, Universidade de São Paulo, São Paulo, 1993.

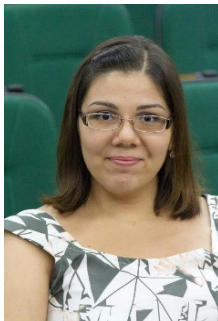
MORAES, J. B. E. de. *A questão do aboutness no texto narrativo de ficção: perspectivas metodológicas para a Ciência da Informação*. 2011. 81f. Tese (Livre-Docência) – Faculdade de Filosofia e Ciências de Marília, Universidade Estadual, Marília, 2011.

MOURA, M. A. Leitor-bibliotecário: interpretação, memória e as contradições da subjetividade. *Perspectivas em Ciência da Informação*, Belo Horizonte, v. 9, n. 2, p. 158-169, jul./dez. 2004.

TÁLAMO, M. G. M. *Comunicação e funcionamento da linguagem*. 1982. Dissertação (Mestrado em Ciências da Comunicação) – Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo, 1982.

TATIT, L. Abordagem do texto. In. FIORIN, J. L. (Org). *Introdução à linguística: I: objetos teóricos*. São Paulo: Contexto, 2007. p. 187-209.

The theoretical model of the Generative Trajectory of Meaning for the preparation of abstracts of scientific texts: a rereading of the informational product abstract



Roberta Caroline Vesú Alves
São Paulo State University
robertavesu@gmail.com



João Batista Ernesto de Moraes
São Paulo State University
jota@marilia.unesp.br

Introduction

The elaboration of abstracts constitutes a procedure of Document Analysis, traditionally focused on scientific texts, aiming at the representation of the content of the document in a succinct text (GUIMARÃES, MORAES, GUARIDO, 2007, MOREIRA GONZÁLEZ, 2004; PINTO MOLINA, 1993). For this reason, it respects the characteristics of this type of text, explained by structural theories of Textual Linguistics (MOREIRO GONZÁLEZ, 2004; VAN DIJK, 2000).

However, it has been verified that the Generative Trajectory of Meaning, derived from Discourse Analysis and Greimas' Semiotics, also cooperates to elucidate some characteristics of the scientific text (FIORIN, 2008; GREIMAS; LANDOWSKI, 1979). Therefore, the objective of this study was to analyze the contribution of the theoretical model of the Generative Trajectory of Meaning for the comprehension of the informational product abstract.

The methodology was an exploratory study with literature review for a qualitative analysis of the information.

It was verified that the Generative Trajectory of Meaning elucidates the mode of expression of scientific knowledge in texts and textual organization, which are important elements to be represented in abstracts.

Abstracts in Document Analysis

Document Analysis presents studies on procedures for treating the content of different documents, however, the studies on the elaboration of abstracts are largely focused on scientific texts.

The Document Analysis procedures, according to Guimarães, Moraes, Guarido (2007), consist of analysis, synthesis and representation for retrieval purposes, through the analytical stage (reading of important parts with metacognitive strategies to identify concepts) and synthetic step (selection, condensation in summaries and representation in indexes).

The representation of content, according to Lara (1993), occurs through intensive condensation, generating documentary products such as abstracts, which resemble the original text and its semiotic system, therefore, similar to the structure and theme of the original text .

The abstract consists of the product of the transformation of the original text into another reduced text about the content of a document, with the function of intermediating the information for the user. (PINTO, 2001; PINTO MOLINA, 1993).

The method of summarizing, according to Pinto (2001), consists of the sophisticated process of reduction over textual content and textual construction, performed cognitively to represent the original text, in addition, it also presents a plan of superficial organization that represents the original text. The superstructure of the original text and the theme (macrostructure) should be transferred to the abstract (MOREIRO GONZÁLEZ, 2004).

The procedures for preparing abstracts, according to Pinto Molina (1993), are complex, but generally follow methods of reading, analysis, synthesis and representation. In this sense, Pinto (2001, p.65, our translation) says that the general process of summarizing comprises,

[...] cognitive activities practiced on linguistic objects through the use of logical tools with a documentary purpose. The result is an abstract/product, a brief representative, intentional, not univocal text and with vocation of synonym. The intrinsic singularity of this analytical/synthetic operative complex, the typological documentary diversity, the personalization of the documentary demands and the variability of the production conditions are, among other aspects, determining factors of the degree of complexity of such operations.

According to Pinto Molina (1993, 2001), the reading for analysis of the original text identifies its fundamental characteristics (form, class and structure of information)

as well as main and secondary themes of the cognitive structure of the document and the important sections, with the essence of the document. Therefore, according to the author, ANSI/NISO Z39.14-1997 (2009) and Simões (2014), the reading observes the parts of the scientific texts that comprise: introduction, objectives, methodology, results and conclusion (IOMRC).

In this sense, reading produces several cognitive processes of reduction and condensation through macro rules, which leads to the global macrostructure (PINTO, 2001; VAN DIJK, 2000). Macro rules show how the act of summarizing cognitively takes place, considering information pertinent to a task, schemes and objectives of the reader to form the macrostructures (GUIMARÃES, 2013).

The synthesis for representation purposes, according to Pinto Molina (1993), composes the information resulting from the analysis in brief, considering the cohesion and coherence, in a way faithful to the original, precise, clear and concise.

The elaboration of abstracts should also consider aboutness, that is, the intrinsic and relatively permanent content, according to Beghtol (1983). Aboutness can be identified through the Generative Trajectory of Meaning (MORAES, 2011). Therefore, we follow with the structural and thematic aspects of this theoretical model and Greimas' Semiotics, important for abstract representation.

Generative Trajectory of Meaning and Semiotics of scientific discourse for the elaboration of abstracts

The Generative Trajectory of Meaning, according to Moraes (2011), contributes to the identification of aboutness revealed by the rules of composition of the narrative and fiction text. This also occurs with other types of texts because of the narrativity. This, according to Fiorin (2008), consists of a minimal narrative, which occurs with the transformation of the state of something (initial state, transformation and final state), which allows the analysis of the structure and meaning of the scientific text, also considering semiotics of the scientific discourse established by Greimas.

The theoretical model of the Generative Trajectory of Meaning demonstrates the levels of production and interpretation of the meaning of the text. Its structural or syntactic and semantic elements are distributed in levels that comprise the deep or fundamental (opposite terms with some common trait), the narrative (phases of manipulation, competence, performance and sanction), and the discursive (themes, figures and trajectories), using isotopes (FIORIN, 2008).

Aboutness can be identified in Document Analysis, from this theoretical model, considering the phases of manipulation and sanction in the narrative, forming the main theme in deep structure (MORAES, 2011). In addition, these phases of the narrative evidence the initial and final state of scientific knowledge, important for the proof of the aboutness of scientific discourse.

According to Greimas and Landowski (1979), scientific discourse can be analyzed through the cognitive and pragmatic dimensions. The first identifies the idea of the transformation of the state of scientific knowledge (initial idea of not-knowing, transformation and final state of knowing, with conjunction with object of value), and the manipulation phase expresses the not-knowing something, and sanction the knowledge reached, to compose the main theme. In this context, cognitive competence corresponds to the knowledge necessary to pursuit of knowledge and performance, the development of the change of state (GREIMAS and LANDOWSKI, 1979).

For this, according to Greimas and Courtés (2008), the sense stated by the researcher on scientific knowledge is expressed in a pragmatic dimension (interior to the cognitive dimension), according to the narrative programs (events perceived by the audience). According to Barros (2008), the narrative programs present the model: $PN=F[S_1 \rightarrow (S_2 \wedge O_v)]$ (F function/do; \rightarrow transformation; S1 subject of doing; S2 subject of state; \wedge conjunction; O_v object of value).

Therefore, one can identify the search of scientific knowledge in a cognitive dimension, through the pragmatic dimension, and the main parts of this type of discourse to compose the abstract. In this sense, it is understood that the elements of scientific research observed in Cervo and Bervian (2005), are related to Greimas' theories as follows:

- Introduction: context that motivated the elaboration of the research, including in general, the theme of the research, problem, hypotheses, thesis, motivations (initial idea of not-knowing, leading to the pursuit of knowledge - manipulation);
- Objective: goals to be reached to prove hypotheses and propositions (idea that guides the transformation of state and the search for knowledge - manipulation aiming at sanction);
- Methodology: means for collecting and analyzing information, techniques for establishing the truth (means by which knowledge will be achieved; knowledge that contributes to the search for knowledge - competence);

- Results: exposes how the collection and analysis of experimental and theoretical information occurred, as well as their relationships (development of change of state or transformation - performance);
- Conclusion: presents the analysis of the information in relation to the proposed objectives, proving or challenging hypotheses, and solving the research problem (final state of knowing - sanction).

To demonstrate the construction of abstracts through these theories, the representation in a structured information abstract was chosen, using the text by Queiroz (2015). Initially, the macrostructure and aboutness were established, observing the phases of manipulation and sanction:

- Manipulation: the initial idea of not-knowing, which leads to the search for knowledge was identified in the research problem and general objective, because the study intends to analyze if the available Assistive Technology resources in multifunction resource classrooms meet the functional demand of students with physical disabilities.
- PN = F [doubt -> (Assistive Technology resources \wedge suitable for students with disabilities)]
- Sanction: Assistive Technology in multifunction resource classrooms such as teaching materials, equipment and furniture partially serves students with physical disabilities, as some students find difficulties in their use and manipulation;
- PN=F[know -> (Assistive Technology resources \wedge partial inadequacy to students with disabilities)].

It was verified in the pragmatic discourse that a partial conjunction with the object occurred in the sanction. Cognitive discourse recognizes the transformation of the initial state of not-knowing (doubt) whether Assistive Technology attends physically disabled students to the final state of knowing that it partially does, which resulted in the aboutness "adequacy versus inadequacy" or, more specifically, "doubt about the adequacy of Assistive Technology resources for disabled students versus partial inadequacy of Assistive Technology resources for students with disabilities".

From this, the structured abstract of Queiroz's text (2015) was elaborated considering the phases of manipulation (introduction and objective), competence (methodology), performance and sanction (conclusion):

Introduction: school inclusion theoretically allows the reception and development of skills and abilities of students with disabilities, but presents difficulties in practice, which

lead to questions of whether the Assistive Technology resources meet the demand of students with physical disabilities. **Objective:** it analyzes whether the Assistive Technology resources of multifunction resource classrooms meet the demand of students with physical disabilities, during the Special Education Service. **Methodology:** triangulation of the qualitative and quantitative information collected through the School Function Assessment to identify the students' functional profile, and Assistive Technology for Education II questionnaires, to analyze the resources used, and semi-structured and non-standardized interviews, applied with seven teachers from the Special Education Service in the Municipal Education Network, working with students with physical disabilities. **Results:** the functional profile of the fifteen students with physical disabilities consists of students between six and fourteen years, attending elementary school, ten patients with Cerebral Palsy and five with Myelomeningocele, who used manual wheelchair to move, Canadian walking stick and three of them had independent walking. In addition, six students used handwriting, six the computer, two the mobile letters and one did not write, and seven performed activities close to the level of their series. The evaluation of teachers on the adequacy of the Assistive Technology resources in multifunction resource classrooms to the needs of students with physical disabilities pointed out some difficulties of mobility and accessibility, inadequate furniture, among other aspects. **Conclusion:** Assistive Technology resources such as teaching materials, equipment and furniture partially served the students with physical disabilities, as some students find difficulties in their use and manipulation.

In view of the above, it is understood that the process of summarizing can also be elucidated by the theoretical elements of the Generative Trajectory of Meaning and Semiotics of the scientific text.

Conclusions

The Generative Trajectory of Meaning and Greimas' Semiotics have contributed to elucidate the way of expression of scientific knowledge in texts, their sense (aboutness) and the underlying structures of textual organization. The elements identified from the proposed theory contributed to the clarification of representation in abstracts, offering a rereading of this informational product in a more conscious way about its structure and thematic.

The scientific knowledge observed through the narrativity (initial state, transformation and final state), pragmatic and cognitive dimensions of the text and the Generative Trajectory of Meaning revealed the textual macrostructure and aboutness, in addition to the important parts for representation in abstracts, that correspond to the superstructure (Introduction, objective, methodology, results, conclusion).

Therefore, it was verified that the theoretical model of the Generative Trajectory of Meaning and Greimas' Semiotics act strategically for understanding the meaning of

the scientific text and its structural aspects, important for understanding the elaboration of the documentary product abstract.

References

The references was made following the ABNT rules.

ANSI/NISO Z39.14-1997(R2009). *Guidelines for Abstracts*. NISO. Baltimore: Maryland,2009.

BARROS, D.L.P.de. *Teoria semiótica do texto*. São Paulo: Ática,2008.

BEGHTOL, C. Bibliographic classification theory and text linguistics. *Journal of Documentation*, London, v.42, n.2, p.84-113, june.1986.

CERVO, A.L.; BERVIAN, P.A. *Metodologia científica*. Pearson: São Paulo,2005.

FIORIN, J.L. *Elementos de análise do discurso*. São Paulo: Contexto,2008.

GREIMAS, A.J.; COURTÉS, J. *Dicionário de Semiótica*. São Paulo: Contexto,2008.

GREIMAS, A.J.; LANDOWSKI, E. *Análise do discurso em Ciências Sociais*. São Paulo: Global,1979.

GUIMARÃES, E. *Texto, discurso e ensino*. São Paulo: Contexto,2013.

GUIMARÃES, J.A.C. O resumo como instrumento para a divulgação e a pesquisa científica. *Revista Brasileira de Educação Especial*, Marília, v.11, n.1, p.3-16, jan./abr.2005.

GUIMARÃES, J.A.C.; MORAES, J.B.E.de; GUARIDO, M.D.M. Análisis documental de contenido de textos narrativos. *Ibersid*, Zaragoza, v.1, p.93-100.2007.

LARA, M.L.G.de. *A representação documentária*. Dissertação (Mestrado em Biblioteconomia e Documentação) – ECA, USP, São Paulo,1993.

MORAES, J.B.E.de. *A questão do aboutness no texto narrativo de ficção*. Tese (Livre-Docência em Linguística e Documentação) – FFC, UNESP, Marília,2011.

MOREIRO GONZÁLEZ, J.A. *El contenido de los documentos textuales*. Gijón: Trea,2004.

PINTO MOLINA, M. *Análisis documental*. Madrid: Eudema,1993.

PINTO, M. *El resumen documental*. Madrid: FGSR,2001.

QUEIROZ, F.M.M.G.de. *Tecnologia assistiva e perfil funcional dos alunos com deficiência física nas salas de recursos multifuncionais*. Dissertação (Mestrado em Educação)–FFC, UNESP,Marília,2015.

SIMÕES, M. da G. de M. *Resumo*. Coimbra: Universidade de Coimbra,2014. Slide da disciplina “Resumo documental” para PPGCI–Unesp.

VAN DIJK, T.A. **El discurso como estructura y proceso**. Barcelona: Gedisa, 2000.

Vocabulary control and its contribution to research: the case of agrarian sciences in Uruguay



Lucía Simón
Universidad de la República
lsimon@csic.edu.uy



Mario Barité
Universidad de la República
mario.barite@fic.edu.uy

Introduction

The explicit policy in Science and Technology (S & T) in Uruguay has had - since 2005 - a growing development reflected in the country's investment in this area, especially since the creation of the National Agency for Research and Innovation (NARI), whose mission is to execute the State's political and strategic guidelines in Research and Innovation (Bértola, et al, 2005). In the context of this Agency, we highlight the creation - in 2007 - of the National System of Researchers (NSI) with the main objective of strengthening and expanding the scientific community. The determination to promote the entities that work with S & T in Uruguay, is not foreign to the needs of the country, so the National Strategic Plan in Science, Technology and Innovation (NSPSTI) establishes priority sectors and nuclei - Human and Animal Health / Agricultural and agroindustrial production make up the list composed of six additional sectors - to be strengthened that must be taken into account by NARI and NSI (PENCTI, 2010).

The present study aims to highlight how scientific research in Agricultural Sciences in Uruguay responds to these guidelines through the study of the scientific production of the highest level researchers in the area identified in the NSI in category III. To reach this categorization, it is necessary to have doctoral degree, original knowledge production, international recognition, creation and management of

research groups and creation of investigation capacity -institutional and training of researchers.

Although NSI provides access to the curriculum vitae - CV - of its researchers, they provide a partial view of the research themes, since the lack of standardization does not clearly allow to evidence the main addressed themes.

In this sense, it is believed that the application of Knowledge Organization techniques such as vocabulary control are appropriate to map the progress of science and disciplines in general since vocabulary control comprises a

set of techniques and procedures applied on language to solve problems of comprehension, ambiguity, scope and relationship between terms that express concepts, based on the idea that natural language, as expressed in texts and documents of the more diverse species generates - including in communities of specialized and more or less solid and closed users - difficulties of interpretation and communication, which are seen as deviations of the methodical and precise discourse (Barité, 2014)

In addition, two capabilities of this tool of essential importance are added. The first, related to its

capacity to improve efficacy and efficiency of information systems (**National Information Standards Organization**, 2010, p. 1), but also of web navigation systems and other systems that are constructed from the use of language to describe reality. In this line, the principle of consistency, understood as "consistency in the application of principles, rules and procedures in relation to the creation and use of conceptual structures or natural language for classification or indexing purposes" (Barité et al., 2013), has a relationship directly proportional to vocabulary control, since it is presumed that the greater vocabulary control, the greater consistency, and therefore, greater efficiency of the information system, in that it facilitates thematic description, search and retrieval (Barité, 2015).

The second - applied in this study - is its

ability to organize conceptual fields, thematic areas and disciplines through the distillation of the concepts most closely associated with each other in the literature, whether they are clearly established and accepted, whether they are in question. Thus, controlled vocabularies offer a fixed state-of-the-art plate of a specialty at a given time and from a particular perspective, considering both the practices, theories and topics already consolidated as well as their points of rupture and conflict. In its evolution, that plate is transformed into a mobile one, exhibiting the changes produced within a thematic field over time (Barité, 2015)

The present study constitutes a first exploratory approximation on the contribution of the tool to the determination of the criteria determined by the S & T policy in Uruguay to evaluate its relevance according to the results that can be observed in the NSI. For the case studied, this can be a particularly important contribution in at least two ways: first, Uruguay's economic growth depends, to a large extent, on agricultural development (Bértola, et al., 2005) and the correct mapping of the research situation in this area suggests that it will result in a better utilization of economic resources. Second, progress in any discipline requires relatively solid state-of-the-art to guide the way forward.

Research in Agricultural Sciences in Uruguay: description based on vocabulary control

The idea that vocabulary control can be considered as a valid tool to demonstrate the state-of-the-art of any discipline and its eventual agreement with the national guidelines in scientific technological matters is based on the application of the basic functions of vocabulary control derived from the comparative analysis of the content of three of the most prestigious standards in the world (*British Standards Institution*, 2005; ***National Information Standards Organization***, 2010; *International Standard Organization*, 2011).

In this paper, we test the hypothesis that vocabulary control can contribute to an overview on the nuclear themes of investigation in Agrarian Sciences of Uruguay.

Data on scientific production were collected in refereed scientific journals from researchers present at NSI Level III corresponding to the Agrarian area. Because the format of the CV does not provide for mandatory allocation of keywords, the articles - or abstracts - were accessed to extract the terms originally entered by the researchers.

Of the 451 articles extracted from the CV, 69.4% had 1,210 keywords. The rest of the articles did not contain this description. The task of vocabulary control over the 1,210 natural language key words was performed in three stages in which the terms were refined to reach a level of detail that evidences the main areas of research. In all three cases, the five basic functions of vocabulary control -restriction, disambiguation, normalization, hierarchization and relationship- with different levels of detail were taken as basis (Barite, 2014).

The first stage basically consisted of the control of equivalences, of synonyms and variants fulfilling the function of normalization. In the second, the restriction function was also applied, whereas, in the latter case, the hierarchy and relation functions were also applied. The function corresponding to disambiguation was used to a lesser extent, since the area itself contributes to disambiguated terms due to its high specificity.

Table I provides the data corresponding to the controlled terms at each stage.

Table I: Stages of vocabulary control

	Start	Stage I	Stage II	Stage III
Keywords	1210	655	147	19

Source: by the authors

Stage I: Control of equivalences, control of synonyms and variants, normalization.

It consisted in the application of the function normalization of vocabulary control, understanding that it comprises the fixation of form and meaning, and the weighted situation of each term in the conceptual structure of belonging, which tends to a greater coincidence between indexing and user searches (Aitchison, Gilchrist & Bawden, 2000). This function was specified in two levels: first, equivalence control was performed on the 1,210 key words that were translated from the English language.

Second, the synonyms and variants were controlled and only those keywords that were considered preferred according to the literary warrant were maintained.

From this stage, 545 terms -474 corresponding to the equivalence control and 71 to the control of synonyms and variants were excluded - which reduced the field of study to a total of 665 terms.

Stage II: Restriction, normalization

It consisted of the application of the restriction function. Barite (2015) mentions four aspects on which restriction works: number of terms, extension of terms, conceptual scope and relationships. In this investigation, the restriction of the term extension was carried out and those syntagmatic keywords with a structure equal or superior to four words were excluded. In case they exceeded that extension and the level of incidence was significant, the expression was normalized to one with smaller number of words.

Once the stage was completed based on the extension, the normalization operation performed in Stage I was performed again. As a result, 415 were excluded because they contained three or more words and 103 were normalized. The final result of this stage was the exclusion of 518 keywords.

Stage III: Hierarchization, relationships and normalization.

At this point the universe of words was reduced to 147. The functions of hierarchization, relation and normalization (Barite, 2015) were applied, which allowed to exclude a total of 128 keywords.

The relationships cover a total of 83 terms that were excluded from the universe of study. The criterion consisted in the identification of authorized and unauthorized terms and the exclusion of the latter was determined by the principle of literary warrant that allowed to establish the existence of the term, the correspondence of the term with a concept recognized by specialists, the representativeness of the term before possible synonyms and the literary warrant that supports it. The hierarchical function was worked from intra-term relationships, in which the situation of synonyms, quasi synonyms and variants; and relationships between terms, which traditionally are established under different hierarchical or associative relationship modalities were

elucidated (Cintra et al., 2002, Aitchison & Clarke, 2004, Naumis, 2007). Only those terms representing associations with relatively high frequencies were recorded. This process excluded 18 terms. As in the previous stages, the terms were again normalized and a total of 128 terms were excluded.

The use of the functions of vocabulary control allowed to move from a universe of 1210 terms to one of 14. Table II shows the two nuclear terms that were determined and two additional columns are added with 12 related terms - at two different hierarchical levels - for each nuclear term. One related term on a second level was recorded three times because it corresponded with that number of different aspects. The fourth column reflects the number of keywords containing the terms.

Table II - Main research topics in Agricultural Sciences of Uruguay

Nuclear Terms	Related Terms	Related Terms	Keywords
<i>Animals</i>			795
	<i>Animal behavior</i>		647
		<i>Reproductive</i>	166
		<i>Productive</i>	98
	<i>Diseases</i>		698
		<i>Cancer</i>	589
		<i>Reproduction</i>	673
		<i>Indicators-Share</i>	66
		<i>Cell behavior</i>	72
		<i>Techniques</i>	54
<i>Land use</i>			803
	<i>Crops</i>		629
		<i>Techniques</i>	305
		<i>Organic Components</i>	398
	<i>Grazing</i>		710
		<i>Techniques</i>	302

Conclusions

The present paper showed that, based on the five typical functions of vocabulary control, even the terms contributed by specialized public require the application of specialized techniques that allow the validation of the terminology of the area to construct more efficiently disciplinary maps that represent the accumulated work in any area of knowledge.

It can be observed that research in Uruguay has two main focuses: animals and land use. The first is based on animal and productive reproductive behavior (milk and meat production) and the diseases associated with reproduction and cancer. To a lesser extent, techniques are provided to combat them. The second focus refers to cultivation and pasturage techniques.

In addition, the usefulness of vocabulary control can be explored for the evaluation of S & T policy in any country. In the case studied, comparing the priority areas established in NSPSTI with the lines carried out by the most prestigious researchers in the country shows the correlation between the two.

It is believed that collaborative work with those who contribute to traditional information systems and those who discuss S & T policy is convenient. A mechanism that allows the researcher to access the most used terms in their area, refined based on the functions of vocabulary control could strengthen both systems.

References

Aitchison, J. & Clarke, D. (2004). The thesaurus : a historical viewpoint, with a look to the future. *Cataloging & Classification Quarterly*, v. 37, n. 3/4, p. 5-21.

Aitchison, J., Gilchrist, A. & Bawden, D. (2000). *Thesaurus construction and use : a practical manual*. Londres: ASLIB.

Barité, M. (2014) El control de vocabulario en la era digital: revisión conceptual. *Scire* v. 20, n.1. Disponible en: <<http://ibersid.eu/ojs/index.php/scire/article/view/4196/3766>> Acceso: 23 may. 2015.

Barité, M. (2015). Control de vocabulario : orígenes, evolución y proyección. *Ciência da Informação*. (Artículo aceptado para su publicación)

Barité, M., et al. (2013). *Diccionario de Organización del Conocimiento : Clasificación, Indización, Terminología*. Montevideo: PRODIC. Disponible en: www.prodic.edu.uy/investigacion/KOD Acceso: 15 de may. 2015

Bértola, L., et al. (2005). *Ciencia, tecnología e innovación en Uruguay: diagnóstico, prospectiva y políticas*. (Serie de Notas de Referencia) S.I: BID.

British Standards Institution (2005-2008). *Structured vocabularies for information retrieval : Guide : BS 8723-1*. London: BSI. Consta de 5 partes: Parte 1: Definitions, symbols and abbreviations (2005). Parte 2: Thesauri (2005). Parte 3: Vocabularies other than thesauri (2007). Parte 4: Interoperability between vocabularies (2007). Parte 5: Formats and protocols for data exchange (2008).

Cintra, A. M. M., et al (2002). *Para entender as linguagens documentárias*. 2a. ed. São Paulo: Polis; APB.

International Standard Organization (2011). International Standard ISO 25.964-1 : Information and Documentation : Thesauri and interoperability with other vocabularies. Geneva: ISO. Consta de 2 partes: Parte 1: Thesauri for information retrieval. Parte 2: Interoperability with other vocabularies.

National Information Standards Organization (2010). Guidelines for the construction, format and management of monolingual controlled vocabularies : an American National Standard ANSI/NISO Z39.19-2005(R2010) developed by the National Information Standards Organization. Bethesda: NISO Press. Disponible en: http://www.niso.org/apps/group_public/download.php/12591/z39-19-2005r2010.pdf Acceso: 15 de may. 2015]

Naumis, C. (2007). *Los tesauros documentales y su aplicación en la información impresa, digital y multimedia*. México: UNAM; Buenos Aires: Alfagrama.

Uruguay. *Presidencia de la República* (2010) Plan Estratégico Nacional en Ciencia y Tecnología. Montevideo: IMPO.

Rural cultural goods: an experience report of the taxonomy construction in the context of the historical farms of São Paulo



Mayara Cristina Bernardino
São Carlos Federal University
mayarabernardino@yahoo.com.br



Maria da Graça Melo Simões
São Carlos Federal University
gsimões@darq.uc.pt



Luciana de Souza Gracioso
São Carlos Federal University
luciana@ufscar.br



Luzia Sigoli Fernandes Costa
São Carlos Federal University
luziasigoli@gmail.com.br

1 Introduction

The scope of our research is in the universe of the historical farms of São Paulo where coffee plantation took place during the 18th and 19th centuries in Brazil. The life and production styles built around these farms were constitutive of the Brazilian nation in different aspects. For this reason, we understand them as information space that requires the attention of Information Science. Our views at the phenomena involved in this field are aimed at identifying resources that allow us to record, preserve and promote access to these cultural goods and inevitably promote the flow of this knowledge. Some of these resources go through actions and interventions in knowledge organization and many of the actions involved in knowledge production in this field require pragmatic approaches, which consider language not as a representation of reality but as the realization of something. However, these theoretical analyzes will not be object of this study, although they have sustained all the reflections made throughout the research process. The objective of this report is to describe the results of the investigations developed to support the elaboration of a taxonomy to be used for representation and retrieval of contents on rural cultural goods, in software created specifically for this purpose, named Virtual Memory. The report is the final result of a master's research funded by Fapesp and is part of a set of research results developed in three projects that also received FAPESP funding, coordinated by different institutions (USP São Carlos – 2004/2008; UNICAMP Campinas – 2007/2012; UFSCar São Carlos – 2012-2014).

2 Theoretical, conceptual and methodological foundations

According to the document published by Unesco, entitled Paris Recommendations, 1964, all goods, whether movable or immovable, which are of cultural importance to a nation, are considered as cultural goods. These categories include artistic works of historical, archaeological, ethnological, specimens (such as flora and fauna) and collections in general. (UNESCO, 1964). As for the term heritage, it would be used to designate the set of cultural goods that become property of a country (FONSECA, 2005). Cultural heritage include monuments, sets of works, notable places (UNESCO, 1972) and immaterial cultural heritage include representations, knowledge, instruments, facts, actions or places that communities or individuals recognize as an integral part of their cultural heritage. This immaterial heritage is fed back by the community itself from generation to generation. According to the document produced by Unesco (Paris Recommendations, version 2003), immaterial cultural heritage is manifested as: a) traditions and oral expressions; b) artistic expressions; c) celebrations, social practices, rituals and festive acts; d) knowledge and practices related to nature and to the universe; e) traditional craft techniques (UNESCO, 2003). On the universe of cultural goods, heritage and memory, a significant body of researchers involved in Information Science has been significantly concerned with these themes, especially in Brazil, and their productions have been concentrated in Enancib working groups (National Meeting of Research in Information Science), namely: WG 09 - Museum, heritage, information, and WG10 - Information and memory. From 2010 to 2013, 217 papers were presented and published, within the context of this conference. The urgency in discussing the theme in the field is also shaped in the conference theme in Enancib 2015: "Information, Memory and Heritage: from document to networks".

In the context of these discussions, we bring the Historic Paulista Farms, understanding them as cultural goods with prospects of being configured as national cultural heritage. From this line of thought, different research groups have extended and applied methodologies to catalog and inventory cultural goods of these farms making them available in a free software named Virtual Memory (VM). The projects developed on the theme and related to us, were:

A) "Memória Virtual de São Carlos" (2004-2008), coordinated by Prof. Dr. José Carlos Maldonado (ICMC/USP - Institute of Mathematical and Computer Sciences, University of São Paulo);

B) "Patrimônio Cultural Rural Paulista: espaço privilegiado para pesquisa, educação e turismo" (2007-2012), coordinated by Prof. Dr. Marcos Tognon (Unicamp - University of Campinas);

C) "Critérios e metodologias para a realização do inventário do Patrimônio Cultural Rural Paulista" (2012-2014), coordinated by Profa. Dr. Luzia Sigoli Fernandes Costa (UFSCar - Federal University of São Carlos).

Our study, which was integrated in project C), generally aimed to elaborate a language instrument, specifically a taxonomy, to be used as support for content indexing in MV software. Different theoretical and methodological orientations subsidized the construction of this knowledge representation and retrieval instrument. Among them, special emphasis is given to the international standards that methodologically guide the construction, such as ANSI/NISO Z39.19-2005, which points out principles and procedures in addition to guarantee international quality.

The development of the proposed taxonomy began in 2010, from the execution of the research project: "Organização do conhecimento no contexto das fazendas históricas do estado de São Paulo: indicações gerais para a construção de linguagens de representação da informação", developed by Bernardino and completed in 2012, as scientific initiation research funded by FAPESP. The main objective of this project was:

A) to construct terminological database related to cultural goods based on the consultation of scientific, literary and technical works related to the domain (based on the theoretical contribution of the literary warrant (BEGTHOL, 1986);

B) develop interviews with field researchers to identify concepts, categories, semantic and pragmatic relations between concepts related to the domain (based on the theoretical contribution of the expert user warrant).

This research resulted in the construction of a terminological base with approximately 1,000 items, resulting from the consultation of literature and specialists.

Afterwards, another research project was developed, entitled "Representação da informação de bens culturais: construindo uma taxonomia no contexto das Fazendas históricas paulistas", also developed by Bernardino, completed in 2015, as a master's degree research carried out at the Graduate Program in Science, Technology and Society, UFSCar, funded by Fapesp. In this project, the central objective was to articulate the 1,000 terms already identified, in a minimal taxonomic structure, to serve as an instrument to support the indexing and retrieval of contents in

the MV. To this end, the research sources were expanded to increase the terminological base and, this time, specific language instruments such as *Thesaurus para acervos museológicos*, and *Tesauro do Folclore e Cultura Popular Brasileira* were used.

Finally, 6,848 terms were generated, initially systematized in 16 terminology bases, categorized according to the consulted source. After a systematic analysis of these concepts, the most representative ones were selected to compose the domain taxonomy, with 3639 terms, hierarchically structured predicting some equivalence relations, which will serve as support for the indexing and retrieval of contents in the MV. Associative relations were not established at this stage of the research, due to timing issues, but future projects are intended to follow up on the development of these relationships. For this stage of the research, in addition to the methodological contributions already mentioned, we use Simões (2008) on the conceptual relations in controlled vocabularies and specifically in the definitions of taxonomy systematized by Aganette, Café e Rocha (2010).

Information Description Pattern (IDP) and Virtual Memory (MV): characteristics and challenges

Information Description Standard (IDP) proposes the description of a variety of typologies of goods, both general and specific data relevant to a given type of good, thus providing a large database interoperable across the different Inventoried goods, enabling the information originally isolated to be used in an integrated manner.

The first version of the IDP was initially elaborated in the project "Virtual Memory of São Carlos", when a prototype of the Virtual Memory software was also created. The second version of this pattern was developed in the public policy project "Patrimônio Cultural Rural Paulista: Espaço Privilegiado para a Pesquisa, Educação e Turismo" and the application, testing and validation phase was executed during the project "Critérios e metodologias para a realização de inventário do Patrimônio Cultural Rural Paulista" completed in 2014. Figure 1 shows the 13 initial fields generated for data description, which are divided into 40 subdivisions in the IDP.

Figure 1 - Fields of the Information Description Pattern.

Grupo	Conteúdo do campo	Atributos do bem patrimonial
G1	Identificação da fazenda	Dados sobre a propriedade, o proprietário e o gestor, endereço e coordenadas geográficas, o tipo de propriedade, a proteção e legislação incidente, imagens e demais informações sobre o contexto histórico-cultural em que se insere a propriedade.
G2	Informações gerais do bem patrimonial	Dados sobre a natureza e tipologia do bem, o registro, o título, o nível de descrição ⁹ (arquitetônico, arquivístico, bibliográfico, bem móvel integrado, iconográfico e natural), localização física com as devidas coordenadas do bem.
G3	Autoria	Dados relacionados com a variedade de tipos de autoria ou responsabilidades atribuídas à criação do bem descrito.
G3	Cadastro de autoria	Dados complementares ao Grupo 2 como nome, pseudônimo, função ou atividade e período de existência.
G5	Produção	Dados sobre o local, ano, versões, edições e responsabilidades concernentes a produção do bem patrimonial.
G6	Descrição	Dados de conteúdo, estilo, resumo, características físicas, dimensões e volumetria, análise estilística, iconográfica e o contexto e espacial e temporal do bem descrito ⁹ e meio de acesso.
G7	Audio-Visual	Dados e materiais produzidos pela pesquisa realizada durante o inventário.
G8	Diagnóstico e Intervenção	Dados sobre o estado de conservação do bem, as intervenções realizadas com os seus respectivos responsáveis.
G9	Disponibilidade, uso e proteção.	Dados sobre a situação de disponibilidade, condição de acesso, uso, reprodução, aproveitamento e proteção do bem patrimonial descrito.
G10	Histórico da procedência	Dados sobre os tipos de transação, valor venal, data de aquisição do bem patrimonial.
G11	Assunto e descritores	Palavras-chave. Termo livre da linguagem natural e termos controlados, retirados de um vocabulário pré-definido, que traduzem o conteúdo temático do bem patrimonial ¹⁰ .
G12	Fonte de informação	Dados sobre as fontes de informação utilizadas na pesquisa histórica do bem patrimonial, independente da sua natureza e tipo de bem patrimonial.
G13	Responsável pela pesquisa	Nome do responsável pela pesquisa histórica, data da coleta de dados e anotações complementares feitas pelo pesquisador.

Source: Nakagawa, Costa and Scarpeline (2010, p.48). (Where the second G3 should be read G4).

Regarding group 11 (Subjects and Descriptors), this research aims to diagnose and suggest the development of a language instrument that would support the indexation of goods that will be recorded in the MV. Although this instrument has been developed, policies regarding its use and feeding still need to be defined. It is interesting to clarify that the denomination of the fields "Subject" and "Descriptors" was established by the interdisciplinary team that developed the Information Description Pattern. Its development has taken into account the fact that users/indexers of the system may not necessarily be librarians or specialists in documentation and archiving. Potentially the system will have records made by owners and employees of the farms themselves. In this context, it was decided to use the category "Subject" for recording free words, natural language, general subjects and "Descriptors" for specific subjects and controlled from the use of the proposed Taxonomy.

IDP was used as basis for the design of the Virtual Memory structure (VM). Free software developed by researchers from the Computer Science course of University of São Paulo (ICMC/USP), São Carlos campus - SP, with the coordination and supervision of Prof. Dr. Elisa Yumi Nakagawa, partners in the projects already mentioned. MV is complete and in the testing phase. The developed taxonomy will be incorporated as a database independent to the Virtual Memory system to be consulted. The initial interface for recording cultural goods can be seen in Figure 2:

Figure 2 - Interface for recording goods in the Virtual Memory

Source: Virtual Memory website available at:
<http://143.107.231.114:8080/memoriavirtual/login.jsf>

It should also be added that other research aimed at the development of policies for indexing, use and evaluation of taxonomy in MV are being elaborated within the scope of the Research Group "Cultural Heritage: Memory, Preservation and Sustainable Management" accredited in CNPQ. Moreover, there is a set of research and investigations being carried out by researchers of this group (GRACIOSO, 2010), focused on the discussion and analysis of the language, in its pragmatic perspective, that is, its use, that try to analyze the movement of meaning of the communication processes reflecting the developments that such an approach might suggest in the configuration of information representation instruments in memory and cultural goods contexts.

Initial structuring of the taxonomy on cultural goods to use in Virtual Memory

The initial methodological structure for terminological organization was based on the general structuring of taxonomies, as suggested by ANSI / NISO Standard Z39.19-2005, but with incorporations of some methodological aspects adapted to the domain specificities. Initially the construction of the taxonomy had the following orientation:

1 Elaboration of taxonomy:

- a) Collection of terms based on literary and specialist warrant;
- b) Organization and systematization of concepts identified and selected in concept bases - ordered alphabetically;
- c) Selection of general categories for conceptual systematization;
- d) Inclusion of terms in suggested conceptual categories;
- e) Establishment of hierarchy and synonymy relations between the terms within the categories;
- f) Delimitation of preferred terms for use in indexing.

Based on the taxonomy construction methods presented by González (2011), in which one can make the reuse of ready-made structures complementing them, in order to save time and effort, we decided to use the hierarchical structure developed by Ferrez and Bianchini (Thesaurus for museological collections, 1987) and feed it with the terms listed and systematized in the terminological base created on the domain of historical farms.

A methodological resource used, however, not indicated in normative guidelines, was the use of colors in the final presentation of the instrument, which indicate the terminological base of origin of the term. Each of the terminology base received a color for its storage, while the terms deriving from these base were incorporated into the taxonomy, these colors were maintained in order to allow a graphic visualization of the origin and the warrants involved in the context of the term. The result can be verified in figure 3, which presents an extract of the developed taxonomy.

Figure 3: Example of terminology design using colors to signal the origin of the terminology base used. (Green: terminological base of interviews with researchers; Yellow: terminological base of FAPESP Partial Report. Projeto de Políticas Públicas Patrimônio Cultural Rural Paulista: Espaço Privilegiado para Pesquisa, Educação e Turismo. The terms in italics, colorless, come from the Thesaurus for Museum collections.

CATALOGO
 CERTIDÃO
 TE2 CERTIDÃO DE BATISMO
 CERTIDÃO DE CASAMENTO
 CERTIDÃO DE NASCIMENTO
 CERTIDÃO DE ÓBITO
 CONVITE
 DECALQUE
 DIÁRIO
DICIONARIO DE EPOCA
 DIPLOMA
DOCUMENTO ADMINISTRATIVO
 TE2 INVENTARIO
 TE3 INVENTARIO PATRIMONIAL
 TE4 INFORMAÇÃO PATRIMONIAL

Source: by the authors

The entire structure of the proposed taxonomy, as well as details on the configuration of the terminological bases are available as an appendix in the paper: BERNARDINO, M. C. *Representação da informação de bens culturais: construindo uma taxonomia no contexto das Fazendas históricas paulistas*. São Carlos: UFSCar, 2015. Dissertação (Mestrado em Ciência, Tecnologia e Sociedade). Programa de Pós-Graduação em Ciência, Tecnologia e Sociedade (PPGCTS), Universidade federal de São Carlos, 2015.

Conclusions

In order to record, systematize, preserve and promote access to cultural goods originating from the historic coffee farms in state of São Paulo, Brazil, over the past 10 years, three large research projects funded FAPESP, involving interdisciplinary and inter-institutional teams were developed. The main results of this research were the development of an Information Description Pattern (IDP) used in free software developed within the scope of these projects, entitled Virtual Memory (VM). With the purpose of promoting, in the scope of knowledge organization, a language instrument that could collaborate with the indexing and retrieval of contents in this system, two research projects (undergraduate and master's) were developed, both funded by FAPESP, which aimed to construct terminological bases related to the domain of historical farms and arrangement of these terms in a taxonomic structure. As a result of these projects, a taxonomy on rural cultural goods has been developed with 3,639 terms, organized systematically, which will be linked to the Virtual Memory. This language will be used by an indefinite variety of subjects, who will be able to participate

in activities of recording cultural goods in this system and will also be able to consult it.

This element, added to the nature of cultural goods, makes the construction of this language instrument challenging and conflicting in terms of its purposes. Even so, we believe that the exercise developed throughout this research can serve the field of knowledge organization, perhaps less by its instrumental result, but more by the problematizations that have materialized with respect to the challenge of restricting language to its representation function. In this view, we are convinced that the spaces of memory (but not only them) demand pragmatic positions of analysis that privilege social actions experienced and constructed through the different uses of language. The image that concludes this article makes us reflect on the limits of our language to represent the lived memories, and it sharpens us to want to continue in the search of understanding on the language as an instrument of action.

Figure 4: Senzala - Fazenda Santa Maria do Monjolinho, São Carlos, SP



Source: by the authors, 2012.

References

The references was made following the ABNT rules

AGANETTE, Elisangela; ALVARENGA, Lídia; SOUZA, Renato Rocha. Elementos constitutivos do conceito de Taxonomia. *Inf. & Soc.: Est., João Pessoa*, v. 20, n. 3, p. 77-93, set./dez. 2010. Disponível em: <<http://www.ies.ufpb.br/ojs2/index.php/ies/article/view/3994>>. Acesso em: 15 de maio 2012.

ANSI/NISO Z39.19-2005: guidelines for the construction, format, and management of monolingual controlled vocabularies. Bethesda: NISO, 2005. 172 p.

BERNARDINO, M. C. *Organização do conhecimento do contexto das fazendas históricas do estado de São Paulo: indicações gerais para a construção de linguagens de representação da informação*. Trabalho de Conclusão de Curso (graduação). Universidade Federal de São Carlos. São Carlos, 2012.

BERNARDINO, M. C. *Representação da informação de bens culturais: construindo uma taxonomia no contexto das Fazendas históricas paulistas*. São Carlos: UFSCar, 2015. Dissertação (Mestrado em Ciência, Tecnologia e Sociedade). Programa de Pós-Graduação em Ciência, Tecnologia e Sociedade (PPGCTS), Universidade federal de São Carlos, 2015.

COELHO, Marisa Colmago. *Tesouro do Folclore e Cultura Popular Brasileira*. Disponível em <http://www.cnfcp.gov.br/tesouro/sistemica.html> Acesso em set. 2014.

FERREZ, H. D.; BIANCHINI, M. H. S. *Thesaurus para acervos museológicos*. Fundação Nacional Pró-Memória, Rio de Janeiro, v. 1, 1987. 86p. (série técnica).

GONZÁLEZ, J. A. M. *Linguagens documentárias e vocabulários semânticos para a web: elementos conceituais*. Salvador: EDUFBA, 2011. 128 p.

GRACIOSO, Luciana de Souza. Parâmetros teóricos para elaboração de instrumentos pragmáticos de representação e organização da informação na Web: considerações preliminares sobre uma possível proposta metodológica. *InCID: Revista de Ciência da Informação e Documentação*, Brasil, v. 1, n. 1, p. 138-158, jul. 2010. ISSN 2178-2075. Disponível em: <<http://www.revistas.usp.br/incid/article/view/42310>>. Acesso em: 24 Mai. 2015. doi:<http://dx.doi.org/10.11606/issn.2178-2075.v1i1p138-158>.

HJØRLAND, B. Domain analysis in Information Science: Eleven approaches – traditional well as innovative. *Journal of Documentation*, v.58, n.4, p.422-462, 2002.

HJØRLAND, B. The classification of Psychology: A case study in the classification of a knowledge field. *Knowledge Organization*, v.25, n.4, p.162-201, 1998.

NAKAGAWA, E. Y.; COSTA, L. S. F.; SCARPINELI, R. *Fazendas históricas paulistas do século XVIII e XIX: premissas teóricas e metodológicas para inventariar bens patrimoniais*. *Resgate*, Campinas, v. 18, n. 20, 2010, p. 37-53. Disponível em: <<http://www.cmu.unicamp.br/seer/index.php/resgate/article/view/305/301>>. Acesso em: 22 out. 2013.

SIMÕES, M. G. M. *Da abstracção à complexidade formal: relações conceituais num tesouro*. Coimbra: Almedina, 2008.

UNESCO, *Recomendação de Paris*, 1964. Recomendação sobre medidas destinadas a proibir e impedir a exportação, a importação e a transferência de propriedade ilícitas de bens culturais. Disponível em: <<http://www.iphan.gov.br/baixaFcdAnexo.do?id=4676>> Acesso: jan. 2015

UNESCO, *Recomendação de Paris*, 1972. Convenção sobre a Proteção do Patrimônio Mundial, Cultural e Natural. Disponível em: <<http://portal.iphan.gov.br/portal/baixaFcdAnexo.do?id=244>> Acesso: jan. 2015

UNESCO, *Recomendação de Paris*, 2003. Convenção para a salvaguarda do patrimônio cultural imaterial. Conferência geral da Organização das Nações Unidas para a Educação, a Ciência e a Cultura doravante denominada “UNESCO” em sua 32ª sessão. Disponível em <<http://portal.iphan.gov.br/portal/baixaFcdAnexo.do?id=271>> Acesso jan. 2015.

Harmonization of CIDOC CRM ontology in the context of archives, libraries and museums



Laís Barbudo Carrasco
São Paulo State University
laiscarrasco@hotmail.com



Silvana Ap. Borsetti Gregório Vidotti
São Paulo State University
vidotti@marilia.unesp.br



Phil Manfred Thaller
Universität zu Köln
manfred.thaller@uni-koeln.de

Introduction

By the fact that Digital Cultural Heritage Repositories, such as Libraries, Archives and Museums, use different metadata standards to describe their information resources, the metadata harmonization from the cultural heritage field is a challenge, because the data models are more designed on the community requirements than on requirements of cross-community interoperability.

In order to integrate information from heterogeneous sources, ontologies as semantic technologies, are already being used and CIDOC Conceptual Reference Model (CRM) is a very prominent ontology used for such purposes.

The **CIDOC CRM** is intended to promote a shared understanding of cultural heritage information by providing a common and extensible semantic framework that any cultural heritage information can be

mapped to. [...] In this way, it can provide the "semantic glue" needed to mediate between different sources of cultural heritage information, such as that published by museums, libraries and archives. (CIDOC CRM)

Metadata are forms used to describe, manage, catalog, and classify documents (information objects). In Libraries, they are traditionally used to describe books, in other words, bibliographic data. In addition, in museums, metadata describes artifacts, paintings, and sculptures. On other hand, in Archive, the metadata are used to describe finding aids, charters, and official documents. Therewith, this treated information can be effectively retrieval. The cultural memory institutions aim to treat their information in order to obtain access and retrieval.

A way to make metadata interoperate is mapping them into an ontology. CIDOC Conceptual Reference Model (CIDOC CRM) is an ontology from the cultural heritage domain, which has been created as a tool for information integration.

Cultural heritage institutions such as museums, archives, or libraries are confronting a crescent necessity to integrate their system. For this reason, to make cultural resources accessible, it is necessary the use of rich metadata structures, capable to cover the variety of material held in these memory institutions.

In this context, ontologies are used as an important tool for achieving information integration, in other words, metadata can be semantically mapped and integrated into an ontology, which has the competence not only to conceptualize specific domains, but also to express their semantics.

The CIDOC CRM arised from the CIDOC Documentation Standards Group in the International Committee for Documentation of the International Council of Museums and CIDOC CRM was accepted as the ISO 21127 in 2006. The purpose of the CRM model is to provide a common language for heterogeneous information systems, and to permit their integration, despite possible semantic and structural incompatibilities. In that way, cultural heritage information can be exchanged and retrieved; and cultural heritage institutions can make their information systems interoperable without having to compromise their specific needs or the current level of precision of their data. (CIDOC CRM)

In summary, CIDOC CRM is a high-level ontology to permit information integration for cultural heritage data and their correlation with museum, library and archive information, which is easily converted to other machine readable formats such as RDF and XML. Possibly the most ambitious application of the CRM is in the

development of integrated query tools, mediation systems and data warehouses. At present, much of the information stored in library catalogues, archival finding aids and museum collection management systems remains isolated. Different information resources normally need to be queried individually, and cross-system links are rare. The ability to combine and integrate information from multiple sources has the potential to add significant value to existing data - facilitating research and enhancing the quality of the user's experience. (ARTUR; CROFTS; LE BOEUF; 2002).

When a user makes a search in a data base he or she doesn't look for a library resource or an archival resource or a museum resource, the user searches for information. In this way, when the user insert a query into the search engine the user does not take into consideration the nature of the resource, that's why it's very important to develop a harmonization of CIDOC CRM ontology in the context of archives, libraries and museums, because it will be able to understand how the CRM entities can be interpreted in the context of archives, libraries and museums and then it will be possible to potentialize the information retrieval.

State of art

Library, Archive, and Museums are repositories of collective memory and they have some functions, for example, the maintenance and conservation of collections; the exposure; and provision of the means for the objects or information retrieval.

Storing, cataloging, classifying and exhibiting objects are part of the relationship process of the man with the world, and they are operations that do not arise with the museum, nor with library, and nor with the archive, but with collections.

Library, Archive and Museum – all the three institutions preserve the memory and organize the access to information. On other hand, according to the institution, the role assigned to the document is different. Traditionally, libraries saves book, archives saves legal documents and museum masterpieces. Thus, each typological material has a way to be treated. In this way, it's needed to differentiate what a term refers to each institution. According to the author Alvares (s.d.) archive, library and museum have the following meanings:

- **Archive:** Ordained storage of documents, created by a person or institution in the course of its activity, and preserved to achieve their goals, aiming the utility that they may offer in the future, therefore, the archival information is information

accumulated by institution to prove or inform, as the institution pursues its purposes/ends, with whom it relates, which rights and duties it represents.

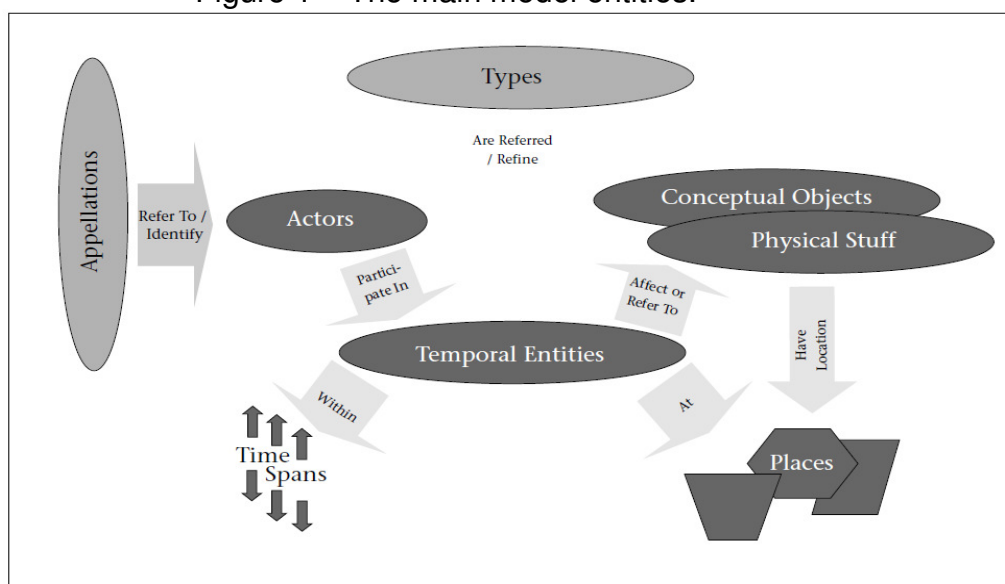
- **Library:** Collection of books, periodicals, audiovisual documents, among others, selected based on their usefulness and arranged to provide access to materials and information, therefore, the organization of the documents in the library reflects the organization given to the area of knowledge (or the subject specifically covered by the documentation center).

- **Museum:** It is a permanent institution, nonprofit organization, in service of society and of its development, open to the public which acquires, conserves, researches, disseminates and exposes the material evidence of people and their environment, for education and delight of society; therefore, the museum, as the instance of the memory representation and social space in which information is considered cultural input, is a fertile ground for the development of studies and actions related to information.

Result

The CRM scope can be defined as all the necessary information for the scientific documentation of cultural heritage collections, in order to enable a broad exchange of information from the area and the integration of heterogeneous sources.

Figure 1 – The main model entities.



Source: Cidoc CRM (DOERR, 2003).

CIDOC CRM contains classes and logical groups of properties. These groups have to do with the notions of participation, structure, location, evaluation and identification, purpose, motivation, usage, and so on. These properties have put temporal entities and with them, the events in a central location (LIMA, 2008).

Table 1 – CIDOC CRM Harmonization in the context of Libraries, Archives and Museums.

CIDOC CRM Entity	Library	Archive	Museum
E1 CRM Entity	CRM Entity about bibliographic material .	CRM Entity about archival material .	CRM Entity about museological material .
E18 Physical Stuff	Literature ("Book") about the Physical Stuff.	Documentation ("Charter") about the Physical Stuff.	Museum object
E28 Conceptual Object	Texts, maps, photos, music, sounds, fairy tales, signs, patterns, symbols, plans, rights, and rules, paper, electronic signals, marks, audio media, paintings, photos, human memory, etc; about circumstances of creation and historical implications of bibliographic material.	Texts, maps, photos, music, sounds, fairy tales, signs, patterns, symbols, plans, rights, and rules, paper, electronic signals, marks, audio media, paintings, photos, human memory, etc; about circumstances of creation and historical implications of archival material.	Texts, maps, photos, music, sounds, fairy tales, signs, patterns, symbols, plans, rights, and rules, paper, electronic signals, marks, audio media, paintings, photos, human memory, etc; about circumstances of creation and historical implications of museum material.
E39 Actor	Author.	Institution.	Artist.
E41 Appellation	All names in the proper sense. Codes or words, meaningless or meaningful, in the script of some group or encoding of an electronic system, used solely to identify a specific instance of some category within a certain context referred to the Library and the bibliographic material.	All names in the proper sense. Codes or words, meaningless or meaningful, in the script of some group or encoding of an electronic system, used solely to identify a specific instance of some category within a certain context referred to the Archive and the archival material.	All names in the proper sense. Codes or words, meaningless or meaningful, in the script of some group or encoding of an electronic system, used solely to identify a specific instance of some category within a certain context referred to the Museum and museological material.
E50 Date	Date of publication.	Date of fund.	Date of creation.
E52 Time-Span	Time-Span of the "Book, Collection".	Time-Span of the "Charter, Fund".	Time-Span of the "Object, Collection".
E53 Place	Place of publication or storage (Library).	Place of provenance or storage (Archive).	Place of creation or storage (Museum).

E55 Type	Typological distinctions of "Book, Collection".	Typological distinctions of "Charter, Fund".	Typological distinctions of "Object, Collection".
----------	---	--	---

Source: Authors.

In this section, it was built from the CIDOC CRM vocabulary a correspondence of terms handled into the cultural heritage institutions – Library, Archive and Museum; because CIDOC CRM entities can have different meanings facing the diverse types of material handled in the three sisters institutions.

Final considerations

We are facing an explosion of varieties of cultural heritage metadata schemas. Cultural Heritage - as the safeguard of the memory - has heterogeneous materials, and for this reason they are safeguarded, protected and conserved in the Libraries, Archives and Museum. In order to treat these digital objects, metadata standards are used. On other hand, how can these three institutions exchange data?

According to the literature, there are many XML metadata mapping to the CIDOC CRM ontology efforts, since this ontology is considered one of the most appropriate models in integration architectures.

In this way, metadata can be mapped into an ontology to provide interoperability of its data and also to achieve information integration. When the different kind of metadata are mapped into an ontology the system can interoperate and the information access is higher as well as their information retrieval.

“An ontology is a specification of a conceptualization” (GRUBER, 1993). More specially, the CIDOC CRM ontology is the specification of the Cultural Heritage conceptualization. CIDOC CRM has an abstract hierarchy.

In summary, the metadata from the three sisters institutions – Library, Archive and Museum – can be mapped into the ontology CIDOC CRM to provide interoperability and information integration of these heterogeneous data from the digital cultural heritage systems.

References

The references was made following the ABNT rules.

ALVARES, Lillian. (s.d.) *Informação em Arquivos, Bibliotecas e Museus*. (Slides) - Faculdade de Ciência da Informação - Universidade de Brasília. Available at: <http://www.alvarestech.com/lillian/Analise/Modulo1/Aula13ABMII.pdf>. Access: May, 2012.

ARTUR, O.; CROFTS, N.; LE BOEUF, P. Elag presentation ontologies. Semantic web and libraries. In: LIBRARY SYSTEMS SEMINAR, 26., 2002, Rome. *Proceedings...* Rome: European Library Automation Group, 2002. Disponível em: <<http://www.ifnet.it/elag2002/papers/pap9.html>>. Acesso em: 18 abr. 2012.

CIDOC CRM. Disponível em:<<http://www.cidoc-crm.org/>>. Acesso em: 20 fev. 2014.

CROFTS et al. (Ed.). *Definition of the Cidoc Conceptual Reference Model*. Cidoc, Icom, 2005. Disponível em: <http://www.cidoc-crm.org/docs/cidoc_crm_version_4.2.pdf>. Acesso em: 28 jan. 2015

CROFTS, N. *Museum informatics: the challenge of integration*. 2004. Tese (Doutorado em Ciências Econômicas e Sociais) – Faculté des Sciences Économiques et sociales, Universidade de Genebra, Genebra, 2004.

DOERR M., The Cidoc CRM: an ontological approach to semantic interoperability of metadata. *AI Magazine*, v. 24, n. 3 p. 75-92, 2003.

DOERR, M.; STEAD, S. *The Cidoc CRM: standard for the integration of cultural information*. Glasgow, Jan. 29, 2008. Disponível em: <http://www.uniurb.it/sbc/ist_bal/seminario/carlo%20meghini/CIDOC%20CRM%20introduction.pdf>. Acesso em: 20 maio 2012.

DOERR et al. The Europeana Data Model (EDM). In: IFLA GENERAL CONFERENCE AND ASSEMBLY, 76., 2010, Gothenburg. *Proceedings...* Gothenburg: Ifla, 2010. Disponível em: <<http://conference.ifla.org/past-wlic/2010/149-doerr-en.pdf>>. Acesso em: 18 fev. 2012.

GRUBER, T. R. *What is an ontology?* Stanford: Stanford University, 2001. Disponível em: <<http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>>. Acesso em: 27 fev. 2012.

GRUBER, T. R. Towards principles of the design of ontologies used for knowledge sharing. In: INTERNATIONAL JOURNAL HUMAN-COMPUTER STUDIES, SUBSTANTIAL REVISION OF PAPER PRESENTED AT THE INTERNATIONAL WORKSHOP ON FORMAL ONTOLOGY, Padova, 1993. *Proceedings...* Padova: Standford University, 1993. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.89.5775>>. Acesso em: 20 jan. 2014.

LE BOEUF, P., Using an ontology-driven system to integrate museum information and library information. In: SYMPOSIUM ON DIGITAL SEMANTIC CONTENT ACROSS CULTURES, 2006, Paris. *Proceedings...* Paris: Louvre, 2006.

LIMA, J. A. O. *Modelo genérico de relacionamentos na organização da informação legislativa e jurídica*. 2008. Tese (Doutorado em Ciência da Informação) – Universidade de Brasília, Brasília, DF, 2008.

Cataloging of interactive resources: analysis of recommendations and practices in international catalogs



Víctor Amor
Investigador independiente
victoramorss@gmail.com



Daniel Martínez-Ávila
São Paulo State University
dmartinezavila@marilia.unesp.br



Rosa San Segundo
Universidad Carlos III de Madrid
rsan@bib.uc3m.es

Introduction

The organization and representation of interactive resources is a growing challenge for both accessing information in libraries and applying traditional standards in new formats. While videogames are increasingly one of the most popular types of resources in libraries for recreational and educational use (Sanford 2008; Levine 2009), the preservation and description of these resources is becoming a major problem for long-term use and memory building. In this sense, several authors have pointed out the importance and challenges of video game preservation in libraries (McDonough et al., 2010, McDonough 2011, Winget 2011 and Watson 2012). As Dan Pinchbeck of University of Portsmouth points out, "it is necessary to catalog and store an enormous amount of information. In particular, games are not usually archived, since they are considered disposable cultural artifacts of little value, although they

represent a really important part of our recent cultural history. Games are one of the most widespread digital media formats on our planet and we must keep them for future generations "(SINC/AG 2009). Henry Lowood, of Stanford University (apud Owens, 2013), also states that the long-term management of this collection goes through the correct management of metadata, ontologies, terminology, etc., that is, aspects related to knowledge organization.

Methodology

In this paper we review the main papers, guides and recommendations in video game cataloging and we analyze some of the main international catalogs in Spain and the United States to determine the practices and degree of adoption of these recommendations in cataloging. The criteria for the selection of the five cases was the prestige of the interactive resource collections: the catalogs of Biblioteca Nacional y de la Comunidad de Madrid, two of the most exhaustive in Spain, and in the United States, the Library of Congress, the University of Oregon and the University of North Carolina, two of the most comprehensive regarding these type of resources.

Videogame cataloging guides and recommendations

Several studies have addressed the theme organization and representation of video games in libraries. James Paul Gee (2007), for example, provides guidelines for the inclusion of video games in library related to education. More recently, Jerome McDonough and others (2010a; 2011) have discussed the possibility of applying the FRBR (Functional Requirements for Bibliographic Records developed by IFLA in 2009) to video games to determine the fields and information needed to retrieve this type of resources in catalogs. Some resources that may provide the necessary information for cataloging include the International Game Developers Association (<http://www.igda.org/>), which provides guidance for cataloging the various persons who appear in the credits of the work, and the website MobyGames. The Authoritative Video Game Database (<http://www.mobygames.com/>) which also provides this data.

As the most notable good practices initiative in the field of librarianship, the project by American Library Association (ALA) Team G (2011) offers information on aspects such as the development of a collection of video games, circulation and access, selection and purchase, purge and, as the most important aspect for the objectives of the present study, the ideal MARC record. Information on this project was

collected after contacting public libraries, school and university libraries in the United States and Canada and experts in the area.

The final report indicates that at least one record must bring information about the title of the video game, the system the game belongs to, the year and publisher, the age classification, the number of possible players, if it has online multiplayer or not, the game synopsis, if using any specific peripheral (eg Kinect, Wii Motion Plus) is necessary, product description, and price. It indicates that the study developer is usually the "author" of the work. In addition to the study, other terms of responsibility to include are director, art director, program director, and voice actors. In cases of libraries that also include consoles in loan, the entry in the catalog should have the following information: the system, items that must be provided along with the console (such as controls, cables, etc.), and price.

Finally, ALA also offers as possibilities for catalogers to use other standards such as Categories for the Description of Works of Art (CDWA), which is mainly used with downloadable and online games, and the Dublin Core general scheme.

Analysis of MARC records in international catalogs

As a first example, we analyzed the cataloging and classification of the game "Age of Empires", by the libraries of Comunidad de Madrid, Spain (see Figure 1).

Field 245 (Title Statement) is headed by the title. Field 250 (Edition Statement), editing could be mentioned, but in the case of the example (which is a collector's edition) it is also noted in field 245. Field 260 (Publication, Distribution, etc. (Imprint)) is noted with the last date on which it has been published/distributed and the publisher (Microsoft) but not the developer (Ensemble Studios). Field 300 (Physical Description) records the support of the interactive resource, whether CD-ROM or other media. Field 300 would not include web resources, these could be in field 856 (Electronic Location and Access). Fields 336 and 337 are noteworthy: Field 336 (Content Type) in MARC helps to specify the content type and includes "Program"; Field 337 (Media Type) includes "electronic". Field 538 (System Details Note) as a field of technical data would be useful when cataloging this type of resource, but it is not used. In the case of the example, field 500 (General Note) includes the characteristics of the computer equipment that allows its use. Field 594 (Local Notes) could be used to include the developer company, other than the publisher, but it is not used either. Field 753 (System Details Access to Computer Files) could be used with these interactive

resources, using \$a for the brand and model of the machine, \$b for the operating system in the case of computer games or mobile phones with their own operating system, etc., however it is not used either.

The field 650 (Subject Added Entry - Topical Term) is not often used in interactive resources, in the case of the example just shows "Computer Games". Field 655 (Index Term - Genre/Form), index term with subfield \$a as Genre/form, would be a deeper cataloging method of this media, but it is not used either. The value shown by Comunidad de Madrid in field 080 (Universal Decimal Classification Number) is 793.7 (086) in the category games, perhaps the number 621.397.48 would also be appropriate given their technological use in anticipation of a more structured and defined classification in the universe of the information of these interactive resources.

Figure 1: MARC fields and situation in the UDC of the interactive resources in Community of Madrid Libraries

	DOCUMENTO	EJEMPLARES (5)	COMENTARIOS	ETIQUETAS
	Visualización MARC ▾			
001	ES-MaRBCM00217257c-2			
003	ES-MaRBCM			
005	20140826:05085200			
008	120620s2000 sp r b spa d			
040	\$aM-BPCM \$bspa \$cM-BPCM			
080	\$a793.7(086)			
090	\$aBPCM -- \$jCDR 793 AGE			
245	00	\$aAge of Empires [\$hRecurso electrónico] \$b: edición para coleccionistas		
260	\$a[USA] : \$bMicrosoft, \$c2000			
300	\$a4 CD-ROM + \$e2 manuales + 2 h. pleg.			
336	\$aPrograma \$2isbdcontent			
337	\$aelectrónico \$2isbdmedia			
500	\$aCaracterísticas del sistema: Windows 95/98/NT 4.0 o superior ; Pentium II 90 Mhz ; 16 ó 32 Mb de RAM (según el juego) ; 100, 200 ó 235 Mb de espacio disponible para la instalación típica (según el juego) ; Lector CD-ROM 4 x o superior ; Super VGA (con 1 Mb de VRAM) ; Tarjeta de sonido compatible con la API ; Módem de 28,8 Kbps para jugar en conexión Internet			
505	0	\$aDisco [1.] Age of Empires : edición gold -- Disco [2.] Age of Empires II : the age of kings -- Disco [3.] Age of Empires II : the conquerors : expansion -- Disco [4.] Age of Empires : banda sonora		
650	4	\$aJuegos de ordenador		
650	4	\$aMúsica para cine		

As a second example, the case of Biblioteca Nacional de España (BNE) is analyzed (see Figure 2). The form of cataloging in the BNE differs from the Community of Madrid in that field 538 (System Details) is used, and not field 500 (General Note)

for the technical requirements necessary to use the resource. Curiously, this note does not show the system to which it belongs. The UDC fields (080) refer to history, Latin and Greek, related to the content of the resource (aboutness) and not to the format.

Figure 2: MARC electronic resource record in Biblioteca Nacional de España

Ver signatura/s		Registro del catálogo	
Palladium [Recurso electrónico]		Cambiar visualización	
Orbaneja García, Luis Miguel			
017	M	17269-2005	bOficina Depósito Legal Madrid
020		84-369-3990-5	
024 8		NIPO	651-05-100-7
080		811.124	(07)
080		811.14	(07)
080		930.85	(37) (07)
080		930.85	(38) (07)
245 00		Palladium h[Recurso electrónico] c[coordinación, Luis Miguel Orbaneja García, Jesús Mª Quiñez Bielsa]	
246 0		iSubtítulo de la etiqueta aCultura y lenguas clásicas. ESO y Bachillerato	
260		Madrid bCNICE, Centro Nacional de Información y Comunicación Educativa c[2005]	
300		1 CD-ROM bson., col. cl2 cm el díptico	
490 0		Serie Recursos educativos multimedia v2005/6	
538		Requisitos del sistema: resolución optimizado a 800 x 600; IE, Netscape (v. 5 ó sup.), Mozilla (1.7 ó sup.); JavaScript habilitado; Flash Player 7	
594		Título tomado de la pantalla de título	
700 1		Orbaneja García, Luis Miguel	

At the Library of Congress in the United States, there is a guide to encode electronic resources (Library of Congress, 2010). This guide is used for computer software (including programs and games) and for "multimedia-oriented computers", systems or online services. The example used in the LOC for cataloging computer games is as follows (Figure 3):

Figure 3: Example of computer game in LOC guide

Computer game

The resource is a game that allows multi-user play online. The type of file is *game*.

Example(s):

World of warcraft, a computer game.

```
Ldr/06: m [type of record: computer file]
Ldr/07: m [bibliographic level: monograph/item]
007:    co#cga||||||| [category of material: electronic resource
           specific material designation: optical disc]
008/26: g [type of computer file: game]
245 00: $a World of warcraft $h [electronic resource].
```

The guide specifies that in field 008/26 - Type of computer file (006/09) - "Code "g" indicates that an electronic resource is a game, intended for recreational or educational use. Generally, games consist of text and software. A video game is

included here." LOC refers to games in general including video games as a recreational or educational resource, with a psrpt of text and software understanding it as a combination of audio and video.

In the example of the LOC (see Figure 4), field 028 (Publisher Number) is used for the editor, which is important because many times the editor is not the developer team, therefore mention to the editor is necessary. In the case of the example Electronic Arts is both editor and developer, while Ambin Interactive is also developer and is not mentioned. Fields 245 and 246 (Varying Form of Title) are used for title and title variation respectively. Field 250 (Edition Statement) mentions the edition (and not in field 245, as in the case of Spain), which in this case corresponds to the system. Field 300 (Physical Description) gives a physical description of the media; Field 538 (System Details) also gives notes on system, console, and command details required for the use of these resources. Field 520 (Summary, etc.) presents a synopsis about the content of the work and 521 (Target Audience Note) the audience to which it is addressed, these two aspects are not contemplated by the Spanish libraries.

Recording does not include classification number (08x). The access point fields include the following information: field 655 in addition to "video games" also includes the system to which it belongs. Field 710 (Established Heading Linking Entry-Corporate Name) also includes the name of the publisher/developer Electronic Arts; Field 753 also includes the name of the system.

Figure 4: MARC electronic resource record in the Library of Congress

Full Record
MARC Tags

[Where to Request](#)

```

000 01570cmm a2200385 a 4500
001 16622114
005 20110124142837.0
007 co cga|||||
008 110124s2008 cau jq g eng d
906 __ |a 0 |b cbc |c copycat |d u |e ncip |f 20 |g y-movingim
955 __ |b qm14 2011-01-24 z-processor
010 __ |a 2011600212
024 1_ |a 014633155471
028 50 |a RVL P RBKE |b Electronic Arts
028 50 |a 1554707 |b Electronic Arts
035 __ |a (OCoLC)ocn227831993
040 __ |a JBL |c JBL |d JFL |d DLC
042 __ |a lccopycat
050 00 |a IN PROCESS
245 00 |a Boom blox.
246 3_ |a Boom blocks
246 1_ |i At head of title: |a Steven Spielberg/EA game
250 __ |a Wii [ed.].
260 __ |a Redwood City, CA : |b Electronic Arts, |c c2008.
300 __ |a 1 computer optical disc : |b sd., col. ; |c 4 3/4 in. + |e 1 booklet (9 p. : ill. ; 18 cm.)
538 __ |a Systems requirements: Wii console; Wii remote.
500 __ |a Title from disc label.
521 8_ |a ESRB rating: E for Everyone; cartoon violence, comic mischief.
520 __ |a Throw, grab, and blast through all sorts of block challenges. Includes over 300 levels set in 5 worlds. Take on single-player or multiplayer co-op and versus gameplay with up to 4 players at once. Tweak any level or build your own and share it with WIIconnect24 (Internet connection required).
650 _0 |a Blocks (Toys) |v Computer games.
655 _0 |a Nintendo Wii video games.
655 _0 |a Video games.
710 2_ |a Electronic Arts (Firm)
710 2_ |a Copyright Collection (Library of Congress) |5 DLC
753 __ |a Nintendo Wii

```

The University of Oregon in the United States has one of the nascent mainstream videogame catalogs (<http://janus.uoregon.edu/search/X>). In this catalog there are also hardware recordings (under the name of material "kit"), which is offered on loan with the software. For example, the "Wii" console displays the following bibliographic record (see Figure 5). In field 100 (Main Entry-Personal Name) the name of the company is included, in the Title 245 (Title Statement) the denomination and in field 500 (General Note) everything it includes: the hardware, the control, the cables, chargers of power, manuals, etc.

Figure 5: MARC hardware record ("kit") at the University of Oregon

```
LEADER 00000nom 2200000 a 4500
099 Game Kit 3
100 Nintendo
245 0 Nintendo Wii
500 Game kit includes: Wii, 2 remotes, 2 nunchuks, sensor bar,
AC adapter and cord, AV cable, component video cable,
eight AA eneloop batteries, battery charger, 3 manuals,
and case.
```

In the example of videogame bibliographic record (see Figure 6) we can observe the use of field 260 (Complex See Reference-Subject) where the developer/publisher (not the distributor) is indicated, also included in 037 (Source of Acquisition) and at Access Point 710 (Established Heading Linking Entry-Corporate Name). The rest of the fields used are very similar to LOC's, with some peculiarities such as the use of field 250 (Edition Statement) which also denotes the system, and the use of Field 500 notes (General Note) instead of 520 (Summary, Etc.) for number of players and the possibility of online multiplayer.

Figure 6: MARC game record at the University of Oregon

```

LEADER 00000cmm 2200409Ka 4500
001 713955177
007 cb cza|||||
008 110421s2011 wau g eng d
024 10 014633098891
037 BLUS 30732|bValve Corp
037 988907|bValve Corp
040 GL9|cGL9|dGL9|dORU
049 [GAME]ORVN
099 GAME 00262 PS3
245 00 Portal 2|h[electronic resource]
246 3 Portal two
250 Playstation 3
260 Bellevue, Wash. :|bValve Corporation,|cc2011
300 1 computer optical disc :|bsd., col. ;|c4 3/4 in. + 1
instruction booklet (9 p. : col. ill. ; 15 cm.)
500 Title from disc surface
500 For 1-2 players (network players, 2 players; network
players co-op, 2). Other features supported: dualshock 3;
headset
500 Accompanying material may vary
520 Set far in the future from the original Portal game, awake
in the Aperture Science Labs as Chell, a former test
subject trying to gain her freedom from the facility and
GLaDOS, the AI controller of the facility. In escaping
your initial confinement, GLaDOS is also reawakened. Set
in the test chambers of the facility, your goal is to make
your way through the game levels using your portal gun to
create temporary passages through solid surfaces, allowing
for creative platforming and multiple possible means of
clearing a level. Puzzle solving at times is needed to
clear sections of levels
521 8 ESRB content rating: E 10+, Everyone 10+
538 System requirements: Playstation 3 game system; 24 MB hard
disk space; supported HD video output: 480p/720P. Optional
online play requires broadband Internet connection
538 Blu-ray disc; Dolby digital
650 0 Artificial intelligence|vComputer games
655 0 Video games
710 2 Valve (Firm)
753 Playstation 3
856 4 |uhttp://libweb.uoregon.edu/scilib/games-info.html|zGame
and Equipment Policies

```

The University of North Carolina also has a catalog that allows searching for "video games", thematic and genre (<http://www.lib.ncsu.edu/catalog/>). For our example, we analyze "Assassins Creed 3" (see Figure 7), which has the added interest that authentic historical facts tells a fictional story, so it included a group of historians for its realization (Rodríguez-Salces, 2012). However, none of these statements of responsibility are indicated in the cataloging. In general, the fields used are similar to the other US libraries', mainly those of the University of Oregon. As a novelty, the use of field 082 (Dewey Decimal Classification Number), which for example 793,932

(Computer fantasy games) is shown, which demonstrates a denotation of the format and not aboutness as in Biblioteca Nacional de España.

Figure 7: MARC game record at the University of North Carolina

Location	Details	Marc Record
		000 02314cmm a2200553Ia 4500
		001 ocn815767846
		003 OCoLC
		005 20121220053204.0
		007 co cga
		008 121102s2012 cau eq g eng d
		024 1 a008888347231
		024 1 a008888397236
		028 5 1 aBLUS-30991 bUbisoft
		028 5 1 aBPSS-306930 bUbisoft
		028 5 1 a347231-DISC bUbisoft
		035 a(OCoLC)815767846 z(OCoLC)815523881
		035 a(OCoLC)815767846
		040 aJBL cJBL dJAI dWAU
		043 an-us---
		049 aNRCC xbw
		082 0 4 a793.932 223
		090 aGV1469.35.A87 bA873 2012
		245 0 0 aAssassin's creed III h[electronic resource].
		246 3 aAssassin's creed 3
		246 3 aAssassin's creed three
		250 aPlayStation 3.
		260 aSan Francisco, Calif. : bUbisoft, cc2012.
		300 al computer optical disc : bsd., col. ; c4 3/4 in. + e1 folded sheet ([10] p. : ill. ; 15 cm.)
		500 aTitle from disc label.
		500 al player; 2-8 competitive network players.
		500 aOther features supported: multiplayer, leaderboards, lobbies/matchmaking, messaging/friend invite in game, voice chat, add-on content, trophies.
		520 a"The American colonies, 1775. As a native American assassin fights to protect his land and his people, he will ignite the flames of a young nation's revolution. Conspire with real historical leaders [and] engage in naval warfare"--Container insert.
		521 8 aESRB rating: M, Mature 17+ (blood, intense violence, sexual themes, strong language).
		538 aSystem requirements: PlayStation 3; 4.5 GB hard drive space; 480p/720p/1080i/1080p HD video output; broadband required; DualShock 3; headset; 3D game; Blu-ray disc.
		596 al
		650 0 aAssassins vComputer games.
		650 0 aQuests (Expeditions) vComputer games.
		650 0 aRevenge vComputer games.
		650 0 aHand-to-hand fighting vComputer games.
		651 0 aUnited States xHistory yRevolution, 1775-1783 vComputer games.
		655 7 aSony Playstation 3 video games. 2NCRS
		655 0 aComputer adventure games.
		655 0 aVideo games.
		710 2 aUbi Soft Entertainment (Firm)
		753 aPlayStation 3
		994 aCO bNRC
		918 a2731203
		909 a20121220
		019 a815523881

In Table 1, a comparison of the information recommended by ALA, which are contemplated by the different analyzed libraries, as well as the MARC fields where they are included, is shown as a synthesis.

Table 1: Information and MARC fields used in the examples of the different libraries

	Comunidad de Madrid	Biblioteca Nacional de España	Library of Congress	University of Oregon	University of North Carolina
Title of the videogame	245	245	245, 246	245, 246	245, 246
System to which it belongs	500	*538	538, 655, 753	250, 538, 753	250, 538, 655, 753
Year of editor/editor	260	260	260	260	260
Classification by ages	-	-	521	521	521
Number of players	-	-	520	500	500
Possibility of online multiplayer	500	-	520	500	500
Game Synopsis	-	-	520	520	520
Necessary Specific Peripherals	500	538	538	538	538
Product Description	300	300	300	300	300
Price	-	-	-	-	-

Conclusions

From the analyzed cases, it is clear, in line with ALA's report (2011), that there is no standardized way of cataloging this type of interactive resources. US libraries generally carry out more complete cataloging than Spanish libraries; no library includes all the information recommended by ALA (such as price information, which is not included in any library). The standards of each library, the way it is cataloged in each one of them and the recent inclusion of this type of interactive resources have made the cataloging very different, as verified with the fact that despite the use of centralized cataloging (copy cataloging), the MARC fields used by each library vary. Sometimes the cataloging depends on the popularity of a title and the speed with which the registration is made by the person who catalogs it. If it is a game that does not exist in a previous cataloging, an original cataloging of it should be done (just as in the gray literature there are certain games without a popularity that condemns them to the "darkness" without having a cataloging and remain as forgotten titles, it is "gray software"). While many sellers and publishers of books provide a cataloging and classification to the library (through the standard ONYX), the pre-cataloging of

interactive resources is in many cases null, so its acceptance as part of a separate collection by libraries is harmed

References

American Library Association. 2011. *2011 ALA Emerging Leader Team G Final Project Page*. Available <http://connect.ala.org/node/146322>.

Gee, James Paul. 2007. *Good video games + good learning: collected essays on video games, learning, and literacy*. New York: Peter Lang International Academic Publishers.

Levine, Jenny. 2009. Gaming and Libraries: Learning Lessons from the Intersections. *Library Technology Reports* July: 11-18.

Library of Congress. 2010. <http://www.loc.gov/marc/ldr06guide.html#game>

McDonough, Jerome, Matthew Kirschenbaum, Doug Reside, Neil Fraistat, and Dennis Jerz. 2010a. Twisty little passages almost all alike: Applying the FRBR Model to a Classic Computer Game. *Digital humanities quarterly* 4 no.2. Available <http://www.digitalhumanities.org/dhq/vol/4/2/000089/000089.html>

McDonough, Jerome, Robert Olendor, Matthew Kirschenbaum, Kari Kraus, Doug Reside, Andrew Phelps, Christopher Eger, Henry Lowood & Susan Rojo. 2010b. *Preserving virtual worlds: Final Report. National Digital Information Infrastructure Program*. Washington, DC: Library of Congress.

McDonough, Jerome P. 2011. Packaging Videogames for Long-Term Preservation: Integrating FRBR and the OAIS Reference Model. *Journal of the American Society for Information Science and Technology* 62 no.1: 171-184.

Owen, Trevor. 2013. *Video Game Preservation at Scale: An Interview with Henry Lowood*. Available <http://blogs.loc.gov/digitalpreservation/2013/02/video-game-preservation-at-scale-an-interview-with-henry-lowood/>.

Rodríguez-Salces, Sergio. 2012. *La introducción de los videojuegos en el aula*. Tesis de master. Universidad de La Rioja. <http://reunir.unir.net/handle/123456789/766>

Sanford, Kathy. 2008. Videogames in the Library? What Is the World Coming To? *School Libraries Worldwide* 14 no.2: 83-88.

SINC/AG. 2009. Desarrollan un emulador para reproducir todos los archivos históricos. Available <http://www.agenciasinc.es/Noticias/Desarrollan-un-emulador-para-reproducir-todos-los-archivos-historicos>.

Watson, David. 2012. Preserving videogames for posterity. *Multimedia Information & Technology* 38 no.2: 30-31.

Winget, Megan A. 2011. Videogame Preservation and Massively Multiplayer Online Role-Playing Games: A Review of the Literature. *Journal of the American Society for Information Science and Technology* 62 no.10: 1869-1883.

Metadata structure for an image base in pathology



Elisabete Gonçalves Souza
Federal Fluminense University
elisabetes.souza@gmail.com



Jóice Cleide Cardoso Ennes de Souza
Federal Fluminense University
E-mail: joicecardoso@id.uff.br



Elan Cardozo Paes de Almeida
Federal Fluminense University
E-mail: elancardozo@id.uff.br

Introduction

The use of images as a biomedical teaching strategy is relevant and contributes to mentor students in the conclusion of diagnoses, researches and studies. Specifically in the scope of undergraduate courses in the biomedical area, the use of visual resources as a teaching methodology has been observed for some time, always aiming to nourish and improve students' knowledge.

A fundamental characteristic of Pathology teaching is the need to recognize macro and microscopic morphological aspects (NEVES, 2008). By getting in contact with images of histopathological slides, the student may reflect on the content delivered in class, allowing them to consolidate all information gathered from the explanation of clinical cases and latest news.

Simultaneously, we observe the democratization of photographic equipment use, which stimulates the production of images in different contexts. The increasing number of images circulating in society, generated and stored in different devices—lpad, smartphones, etc—, leads us to question about the retrieval of a given image from the already existing access points. Broadening this questioning to the scope of the images designed for didactic use, we ask ourselves: what are the metadata needed for the description of digital images of histopathological slides for the purposes of information organization and retrieval?

This article reports the conclusion of the first stage of the project *"Organização de imagens de lâminas histopatológicas: abordagem teórico-metodológica visando à recuperação da informação"* and aims to identify metadata for the description of digital images of macroscopic lesions and histopathological slides produced within the scope of Departamento de Ciências Básicas - FCB (Basic Sciences Department), of Campus Universitário de Nova Friburgo (Nova Friburgo University Campus), Universidade Federal Fluminense (UFF) (Fluminense Federal University) ; and to analyze and contextualize the metadata for their inclusion in the Dublin Core structure, in order to generate the metadata model for the anatomical pathology database to be developed.

Rationale

The Pathology discipline of FCB/UFF has made available to the students of the courses of Speech Therapy, Dentistry and Biomedicine approximately 200 images of

histopathological slides for study in the last four years. These images are currently organized according to the layout of the programmatic contents, presenting only a descriptive caption, thus not offering any form of retrieval to the user. The organization of this collection by the international standards for the description of digital resources will allow the retrieval and access of the images and of all information associated with them (type of lesion, organ, diagnosis, etc.), in addition to favoring the sharing and interoperability of the future image base with other similar information systems.

The study is relevant since the images are important didactic resources for the expositive and laboratory classes, for they make it possible for students to become responsible for their own learning by performing independent solutions (self-instruction) of clinical cases by accessing the images remotely in the discipline's webpage. This drill contributes to the perfection of knowledge beyond the limits of space and time, enriching learning and making it more dynamic through the use of current technologies, such as the creation of an image base.

The literature review has revealed the shortage of Brazilian Portuguese papers that aim to launch a metadata standard for histopathological slide images. We highlight the study by Carrare et al (2006), who developed a digital environment named Virtual Library of Medicine Images (VLMI) for the organization and retrieval of health images with a focus on education. On the metadata, the authors clarify that "a predefined protocol was used with data belonging to the digital images and information pertinent to the standards of records of the health system" (CARRARE et al., 2006, p.203), not explaining the adopted standard.

There are three basic types of metadata for information organization, representation and retrieval: a) textual metadata, used in cataloging, whose main objective is to identify and describe informational resources, the most appropriate standard being AACR2; b) the format of simple, unstructured metadata, extracted automatically by robots, such as meta tags; and c) the structured metadata that supports search engines, helping them to discover the information present in the web environment, such as MARC 21 and Dublin Core standards (ALVES, 2005). In addition to these, we have administrative, technical and preservation metadata, as shown in the figure below:

Figure 1 - Metadata types

TYPES	DEFINITION	EXAMPLES
Administrative	Metadata used in the administration of information resources, since its creation and system introduction	<ul style="list-style-type: none"> • Acquisition of information • Rights of reproduction • Selection requirements for scanning
For description	Metadata used to describe information resources	<ul style="list-style-type: none"> • Cataloging records • Specialized indices
For preservation	Metadata used for the preservation of information resources	<ul style="list-style-type: none"> • Indicate the physical conditions of the resource
Technical	Metadata used in administration to know the functions of a system	<ul style="list-style-type: none"> • Hardware and Software • Safety data • Documentation
For use	Metadata relating to the level and type of use of an information resource	<ul style="list-style-type: none"> • View records • Reuse and versions (summary)

Source: LOURENÇO (2007)

In the setting of image studies, we rely on DICOM (Digital Imaging and Communications in Medicine) metadata standard used worldwide for the description of diagnostic images (MACHADO, 2002). According to Hasegawa and Aires (2007, p. 50), the hegemony of DICOM standard is related to the sale of the equipment, being "a standard published by the National Electrical Manufacturers Association (NEMA) which counts with more than 50 factories supplying medical equipment, in addition to committees and working groups." That is, the application of DICOM standard in applications for medical image organization is directly related to the use of such equipment.

DICOM standard's special feature is bringing as metadata the patient name associated with the image and the medical record number. However, Carro (2003, p.77) points out that, despite making this metadata available, DICOM "does not have enough semantic content for indexing and searching in generic medical image contexts, such as the Web."

In contrast, there is Dublin Core (DC), an international standard for digital objects in general, which is open and available for free. In addition to semantic interoperability, DC metadata is customizable and can be tailored to the needs of users when describing their resource. Its use in the scope of the description of biomedical images is reasonable, since the use of textual metadata increases the speed of image search, making it more efficient and precise. Thus, the similarities between the images and their respective diagnoses will be optimized, since instead of comparing the

images, the user will access all their textual metadata, including the morphological diagnosis with which they relate.

Based on the DC model, Carro (2003) presents the MediSeek class-based metadata model for description and retrieval of web images, favoring specialization and its implementation in the Resource Description Framework (RDF) metadata architecture, markup language that, using XML (eXtensible Markup Language), describes content in a structured way, allowing coding, interchange and processing of metadata on the web.

Aires, Matos and Hasegawa (2013) present the use of the DC standard for indexing medical images for purposes of data storage, organization and retrieval. According to the authors, the proposed standard also serves as the basis for the construction of the database, allowing the relationship between previously defined images and metadata.

Methodology

From the exploratory and descriptive analysis of the existing literature on the application of the Dublin Core standard in the description of biomedical images, a prototype was elaborated with the metadata to be applied in the description of histopathological slides images.

According to Marcondes (2005, p.96), in addition to describing resources, the use of metadata in the context of the web allows "to accommodate computers and special programs, robots and software agents, so that they understand the metadata associated with the documents and can retrieve them." This gives users access to a range of structured information, which they can analyze and evaluate their relevance, as well as manipulate, use and reuse them according to their needs.

The main characteristic of this proposed standard is to allow the specificity in the description of an image by inserting new metadata that contribute to the full description of the resource in a contextualized way, considering its use for didactic purposes.

The DC standard was used for the selection of the metadata to be applied in the description of the histopathological slide images, because it is a description standard for simple, objective and self-explanatory digital resources, thus allowing the authors of the resources themselves to describe them when publishing them.

Composed of fifteen fields, in addition to simplicity, DC has other important characteristics: *flexibility*: the elements that compose it are optional, and can be displayed randomly and, if needed, they can be repeated or modified by qualifiers; *extensibility*: allows diverse communities in distinct areas to use the DC standard by exchanging information and gaining access to it; and *semantic interoperability*: different models of description interfere in the ability of searches between areas, so the application of a common model increases the possibility of interoperability between these areas.

In the process of creating the metadata standard for the description of macroscopic lesions and histopathological slides, we investigated repositories with image collections where we analyzed the structure and content in order to adapt the observed patterns to the image base proposed for Pathology. This was necessary due to the specificity of this tool for the study and definition of morphological diagnosis by image, taking into account that the visual content of an image is insufficient for its interpretation, and it is necessary to describe it fully, since one has access to the "true expressiveness of an image" (CARRO, 2003, p. 15) only when it's accompanied with consistent description of textual information, therefore having the means to retrieve it.

To define the model, tests were performed to adapt the DC metadata to the complexity of the information contained in the biomedical images.

Analysis of results

Metadata standards such as DC have a more formal and detailed description structure, based on specialized norms and codes of a particular field, such as those from documentary sciences (cataloging codes, controlled languages, etc.), which displays information and its content description more densely. This enables a more accurate and contextualized retrieval of resources, such as anatomopathological images, since the base provides fields related to diagnostic information, among others.

In a digital library, such as an image base, metadata standards include descriptive metadata which describes the informational resource (author, title, subject, etc.); the structural ones used for visual structuring of the digital library; and the administrative ones, that control access dates to the information resource, copyright, etc. (LOURENÇO, 2007).

In this direction, we propose the metadata scheme below for the identification and description of the anatomopathological images produced within the scope of the General Pathology discipline of UFF.

The elements were organized in: *Intellectual Property* (Figure 2): Creator, Collaborator, Pathologist, Editor, Rights; *Instantiation* (Figure 3): Creation date, Release date, Publication date, Image base identifier, Image identifier, Sample format / Enlargement / Coloring, Size / Resolution, File type, Language; and *Content* (Figure 4): Title, Image type, Description, Type of lesion, Organ, Morphological diagnosis, Descriptors Relationship, Form of acquisition, Responsible for acquisition, Source.

Figure 2 - Metadata for the description of Intellectual Property Elements

ELEMENTS DUBLIN CORE	NOMENCLATURE	DESCRIPTION	DESCRIPTIVE GUIDELINES
dc.contributor.author	Creator	Name of the creator of the resource	Entry according to AACR2, entry of people names
dc.contributor.advisor	Collaborator	Name of the person responsible for editing the image.	Entry according to AACR2, entry of people names
dc.publisher	Publisher	Name of the institution responsible for the publication of the image.	Entry according to AACR2, entity name. In this specific case, the name of the University / Department
dc.rights	Rights	Information on usage rights.	Authorized use for didactic purposes with reference to the source.

Source: by the authors

For the description of the metadata referring to authorities (people, collaborators, entities, etc.)—Figure 2—the chosen standard was AACR2 (Anglo-American Cataloging Rules), for it is the most used standard by information systems

in Brazil. AACR2 has two specific chapters (Chapter 22 and Chapter 24) for standardizing headers for names of persons and entities.

Figure 3 - Metadata for the description of Instantiation Elements

ELEMENTS DUBLIN CORE	NOMENCLATURE	DESCRIPTION	DESCRIPTIVE GUIDELINES
<i>dc.date.created</i>	Creation date	Resource creation date	Describe: month / day / year
<i>dc.date.available</i>	Release date	Release date of the resource for viewing in the image bank	Describe: month / day / year
<i>dc.date.issued</i>	Publication date	Date of publication of the resource in the image database.	Describe: month / day / year
<i>dc.identifier.banco.uri</i>	Image identifier	Image bank URI	Copy link indicating resource URI
<i>dc.identifier.imagem.uri</i>	Image identifier	Resource URI in image bank	Copy link indicating resource URI
<i>dc.format</i>	Image identifier Sample Format / Magnification / Coloring	Physical description image	Resource description Example: 1 slide. Photo on optical microscopy. Coloring HEHE – 4x
<i>dc.format.extent</i>	Size / Resolution	File size and resolution	Describe size and resolution in dpi. Example: 7 MB – 72 dpi
<i>dc.format.mimetype</i>	File type	Media type and extension	Describe the media type and file extension. Ex.: image/jpeg
<i>dc.language.iso</i>	Language	Language of resource	Describe the language. Example: Por_Br, eng_US

Source: by the authors

Simulation exercises indicated that adjustments were necessary and some fields were repeated to meet the integrity of the description, such as the fields *dc.identifier.banco.uri* and *dc.identifier.imagem.uri*. (Figure 3), whose metadata guarantees access to the resource and the site where the images are stored. As we mentioned before, there is still no image base, but a project to develop it, and the architecture of its metadata being proposed in this study.

Figure 4 - Metadata for the description of Content Elements

ELEMENTS DUBLIN CORE	NOMENCLATURE	DESCRIPTION	DESCRIPTIVE GUIDELINES
dc.title	Title	Resource Title	Describes as it appears on the slide legend
dc.type	Image type	Nature or gender of the resource	Describe the type of resource: Example: Photography
dc.description.abstract	description	Resource Information	Find information in the resource caption
dc.description.lesaotype	Type of lesion	Type of lesion identified	Use the DeCs glossary to describe the term
dc.description.orgao	Organ	Region of the body where the histopathological cut was done	Look for this image in the image caption.
dc.description.diagnostico	Morphological diagnostic	Diagnostic results	Describe the result of the diagnosis as written by the pathologist
dc.contributor.advisor	Patologista	Name of the person responsible for the diagnosis	Describe according to AACR2, entry of people's names
dc.subject	Subject Descriptor	Main subject identified in the resource	Describe in Portuguese, according to the DeCs
dc.subject	Subject Descriptor	Assunto principal identificado no recurso	Describe in Portuguese, according to the DeCs
dc.subject	Subject Descriptor	Another subject matter identified In resource	Describe in Portuguese, according to the DeCs
dc.subject	Keyword	Another subject matter identified In resource	Describe in English, according to the DeCs
dc.relation	Relationship	Indicate sources related to the subject	Includes link to which the subject refers. Example: Articles, papers and other documents
dc.description.origem	Acquisition form	Indicate the form of acquisition	Describe how the resource was acquired. Exemple: donation
dc.description.origin	Responsible for the acquisition	Person or institution responsible for acquiring the resource	Describe according to AACR2, entry of people's names; entity names
dc.source	Source	Type of equipment	Describe the type of equipment that generated the image

Source: by the authors

The concern with consistency in the search and retrieval process led us to add the fields `dc.description.lesaotype`, `dc.description.orgao` and `dc.description.diagnostics` (Figure 4). These fields are particular to the biomedical image and relate, respectively, to the types of lesions, to the region of the body where the macroscopic image was taken and histopathological cut, and to the morphological diagnosis that will assist the clinician in the definition of the final diagnosis. The fields Lesion Type, Organ and Morphological Diagnostic are metadata for the image base

management and will act as "keys" for the first classification cut, contributing to the systematization of its content.

For the indexing of pathological images—dc.subject field—the trilingual controlled vocabulary DeCs (Descriptors in Health Sciences) was chosen because it is considered an authority in the area and easily accessible by users. The DeCs is an initiative by BIREME (Latin American and Caribbean Center on Health Sciences Information) and was prepared from the MeSH (Medical Subject Headings). It aims to be a unique indexing language for health information records and to be an intermediation language for information retrieval of the sources available in the Virtual Health Library (BIREME, 2015). There are 32,160 hierarchically organized descriptors, of which 4,622 are exclusive to DeCs.

In Health, the control of natural language phenomena (synonymy and homonymy) contributes to information sharing across institutions. For this, it is necessary to elaborate artificial languages taking into account the documentation, the user and the information policies. Documentary languages are constructed "for information indexing, storage and retrieval and correspond to symbol systems intended to 'translate' the content of the documents" (Catherine et al., 2002, p.33).

In the context of information organization and representation, textual metadata (descriptive and thematic) guarantee consistency in information search and retrieval processes, and other metadata sets, such as structural and management metadata, will maintain interoperability and system usability, allowing users to share and use the information that is made available to them.

Conclusions

The proposal to create a metadata model for the anatomopathological image database confirms the need to think of tools for the storage of biomedical imaging resources that are accessible to users (professors, pathologists and students) so that when depositing their images they can describe content in an objective and simple way.

The appropriate description of biomedical images will allow the pathologist to retrieve different samples to make comparisons and to verify their morphological diagnosis. The use of a database structure, such as DC, will facilitate information retrieval through authorized descriptors, as well as other data that guarantee the authenticity of its contents: name of the creator, base curator, etc., scientific authorities

to which the images are related, indispensable information for the data recording deposited there.

For the web setting, the DC metadata standard presents itself as a reliable and consolidated alternative for the description of resources, such as image documents, whose level of description is more specific and detailed. As we have seen, one advantage of its application is the set of descriptive elements, which can be divided into two levels of structure and specificity: simple, with fifteen elements of description, and qualified, which included additional elements of description and indication of qualifiers, which Identify schemas for the value of the element, such as the use of documentary languages, such as the DeCS thesaurus, for the subject element and the AACR2 cataloging code for elements such as "author" and "advisor".

All the description elements, including the additional elements, refinement elements and schemata, can be repeated, which allowed thinking of a descriptive structure rich in biomedical information, according to the needs of the pathology studies for the image base.

On top of that, the images will be properly organized and retrieved, with the possibility of increasing the pathological content retrieval on the Web from the interoperable property of the Dublin Core standard and its use in image bases. The application of the metadata standard guarantees a normalized description and, as a consequence, a quality representation that will facilitate information exchange, interoperability across systems and, in particular, information retrieval. This fact promotes the increase of case studies in General Pathology learning, dynamizing the teaching-learning process by undergraduate students.

Research carried out with financial assistance from Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq (National Council for Scientific and Technological Development)

References

The references was made following the ABNT rules.

AIRES, J. P.; MATOS, S.N.; HASEGAWA, F.M. Usando o padrão Dublin Core para definição de metadados de imagens médicas. *Revista Eletrônica FAFIT/FACIC*. Itararé, SP: Faculdades Integradas de Itararé – FAFIT-FACIC, v. 4, n. 2, p. 01-12, jul./dez., 2013.

ALVES, R. C. V. *Web semântica: uma análise focada no uso de metadados*. 2005. 180 f. Dissertação (Mestrado em Ciência da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2005.

BIREME. *DeCs – Descritores em Ciência da Saúde*. Sobre o DeCs, 2015. Disponível em: <<http://decs.bvs.br/P/decsweb2015.htm>>. Acesso em: 17 maio 2015.

CARRARE, A.P. et. al. Uma proposta para gerenciamento e preservação de imagens em medicina na EPM/Unifesp. *Ciência da Informação*, Brasília, v. 35, n. 3, p. 201-208, set./dez. 2006.

CARRO, S.A. *Um modelo de metadados para a indexação e recuperação de imagens médicas na web*. Dissertação (Mestrado em Ciência da Computação). Porto Alegre, 2003.

CINTRA, A. M. M. et.al. *Para entender as linguagens documentárias*. 2. ed. São Paulo: Polis, 2002.

HASEGAWA, F.M.; AIRES, J. P. Proposta de um Padrão de Metadados Para Imagens Médicas. *ERI - Escola Regional de Informática*, Guarapuava, PR, 2007.

LOURENÇO, C.A. Metadados: o grande desafio na organização da web. *Informação & Sociedade*, João Pessoa, v.17, n.1, p.65-72, jan./abr., 2007.

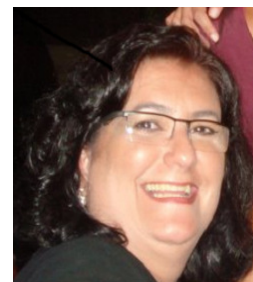
MACHADO, M. S. *Modelo de metadados para armazenamento de recuperação de imagens estáticas no formato DICOM*. PPGC da UFRGS, 2002. Dissertação de mestrado Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Computação.

MARCONDES, C.H. Metadados: descrição e recuperação na Web In: _____ et al. (Org.). *Bibliotecas digitais: saberes e práticas*. Salvador, BA: EDUFBA; Brasília; IBICT, 2005. p. 77-143.

NEVES, F.B.C.S. et al. Impacto da Introdução de Mídia Eletrônica num Curso de Patologia Geral. *Revista Brasileira de Educação Médica*, v. 32, n. 4, p. 431-436, out./dec.2008.

UNIVERSIDADE FEDERAL FLUMINENSE. Faculdade de Odontologia de Nova Friburgo. *O que é patologia*. 2015. Disponível em: <<http://patogeralpunf.wix.com/punfuff#!sobre/c1t44>>. Acesso em: 17 maio 2015.

Indexing through images: accessibility via OPACs



Rita de Cássia do Vale Caribé
University of Brasília
rita.caribe@gmail.com

Marcílio de Brito
University of Brasília
mdebrito@unb.br

Introduction

The postmodern society is characterized by an effort of compulsive and obsessive modernization, by complexity and uncertainty. In this context, information has a relevant role, as it allows individuals to be able to rely on quality information to satisfy their information needs, whether for professional practice, whether for the exercise of citizenship or for their daily activities.

The process of knowledge generation is scattered among a wide variety of institutions and individuals that systematize information in order to facilitate its retrieval and the primary document location when necessary. In a privileged place are the libraries, social units characterized by the organization of sets of informational objects aiming at their retrieval and dissemination in syntony and synchrony with the effective and potential demands of their users.

In the information organization process, specifically at the description stage of the document content, language is used to represent the concepts treated in it.

The issue of content representation becomes complex when dealing with unusual languages, especially those used by individuals with special needs or disabilities.

It is also understood that the nature of the collections is as diverse as the nature of the documents, thus, by solving the problem of access to the document, the document-user communication process is established by the very nature of their affinities. The key aspect is to put the information in contact with its user. Assuming that the library catalog is the main gate through which one searches for documents in a collection, we ask ourselves: how to give access to the library catalogs?

Content description in the traditional sense

Traditionally, the content analysis process identifies the subjects and the form under which they were treated, their temporal and spatial contiguity among other relevant aspects to facilitate the location of the searched document.

The operation named indexing is the act of identifying and describing the content of a document with terms representative of its subjects and which constitute an indexing language" (ABNT, 1992, p. 2). UNISIST (1981) complements it by specifying that indexing makes it possible to retrieve the concepts contained in documents that can be represented by terms selected from natural language (keywords), from a thesaurus (descriptors) or from symbols (classification number taken from a bibliographic classification system, such as DDC or UDC).

In the indexing process, the concepts are extracted from the document through analysis, therefore, in natural language, rich in meanings and figures, such as metaphors, poetic licenses, among others. Natural language is characterized by connotation.

The operation that allows the use of indexing keywords is to transform concepts into terms, the same ones found in thesauri, lists of subject headings, classification systems, etc.

Linguistic approaches were fruitful (BRITO, 1992) and have proved more recently, with the progress of technology, the importance of this procedure. Nevertheless, simplified metadata methods, such as folksonomy tags, propagate considerably in information storage and retrieval systems. The reason for this is the simplicity of the implementation, although the recall and relevancy effects are known in the retrieval.

Not everything, therefore is bad and not everything is efficient at all. Even the most sophisticated linguistic models, from parsers of the full text, confront performance deviations due to the great flexibility of the laws governing the discourse produced by the communities, in contrast to the rigidity of their language grammars.

In this evolution process of the written information treatment tools, there has been a relative evolution in the recognition of form (morphology) and syntactic relations inter and intra discourse units. Systems based mainly on statistics took advantage of this advance attributed to the linguistic methods and the appearance of the ontological systems and the semantic web were observed.

The use of these simplified, hybrid and statistical methods, as opposed to those proposed by linguistic methods, gives the user the competence to identify the relevance of the results. The imprecision of the systems and the uncertainty in the completeness of the search are passively accepted as normal.

The posed problematic is that technological moment asks for new solutions without changing the balance of this cost-benefit relation. To what extent is it possible to modify the production chain (information indexing), without undermining the current documentary institutions, and at the same time increase user satisfaction?

Marketing studies (ASHTON, 2015) demonstrate the advantages of using images and raise important considerations about the information media. Being able to read is essential, but not everyone can read in the same way. The recorded information is not only written information, as image has gained space in this communication process.

Semiology shows that the image is a complete communication language, and in this matter, the "linguistic" competences of the population, widely consumer of images, is often superior to the competences of written language. The communication laws and codes that govern image are therefore more understandable to the general public than the traditional written codes.

Visual reading surpasses the reading performance of written messages with greater precision and by a wider public:

- Brains consume 1/10 second to perceive a landscape image as opposed to the 60 seconds needed to read the description of that same scene (200-250 words).

- Imagine yourself traveling at 80 km/h and reading the following text, "Warning: area subject to rocks falling caused by instability in the slope or by men



working ahead", instead of this image.

- Including videos in Twitter messages increases Retweets by 28%, and tweets that include photos reach 35% more Retweets (ROGERS, 2014);

- A study published by Social Bakers (2014) showed that photos represent 87% of the posts shared on 30,000 pages, while any other type of content does not reach more than 4% of the shares.

Introducing a new concept of OPAC

Given the evidence of this (human) preference for the use of images in communication systems, one wonders why libraries would not adopt this form of communication? After all, communication is part of their cultural values linked to citizen actions and social inclusion, favoring access to information and documentation as part of human rights universality? Would not this be a natural evolution path for today's systems?

These questions evolve in several ways, on the one hand considering the users' characteristics, and on the other hand, studying the technical and functional adaptations of the libraries' bibliographic processes, products and services.

With the use of information technologies, the so-called Online Public Access Catalog (OPAC) currently dominates the library scenario. The literate population accesses the OPACs through interfaces essentially based on written language. The user communicates with the library's collection through menus, alphabetical lists, or Google-style search fields. Although less frequent, children's audiences have graphical interfaces adapted to small collections. At the margin of this process are the minority communities such as the illiterate, both digital and functional, the visually and hearing impaired, the autistic, the elderly and others. If on the one hand the visually impaired have software that make the audio-reading of computer screens, few tools are available to compensate other minorities' barriers.

The reformulation of the current concept of catalog, based on written language, for a new concept of visual/image catalog claims for responses on the emergence of

new technologies, but it is confronted with other issues regarding the epistemology of catalogs, the characteristics and properties of images.

Linguistic references for information processing

The information searching process leads to the verification of a certainly complex relationship among writing, communicating and describing the language to express the thought. In fact, what one wishes is to be able to represent materialized knowledge, through language acts, in a text. This representation passes through the application of methods able to convey intrinsic properties to the language and its grammar. The research that is undertaken here consists, essentially, of transposing documentary content into a meta-language using images as representatives of this controlled (meta) language.

In France, in the 1960s, linguists focused on descriptive research. Problems previously limited were treated from the point of view of "morpho-semantic fields" or from the lexicological perspective dissociated from morphology, working on circumscribed corpus: a semantic description of the vocabulary of housing, domestic animals, among others. Little by little, the physical properties of objects, denoted by the studied words, is used, a certain progress is made from the examination of the codes or documentary languages destined to the retrieval of the objects on the texts of archaeological origins. In the research on semantic theory, especially, the linguistics research on information representation (BARTHES, 1967; GREIMAS, 1966, 1970, 1973, 1976; GREIMAS, s.d.; GUIRAUD, 1966; MOUNIN, 1965), which stemmed the research on the representation of natural language and in particular to its automatic processing, considering the numerous components that intervene in this process - morphological, lexical, syntactic, semantic, and logical - as well as the relations and dependencies among the simple and complex elements of discourse.

It is known, however, that somewhere in the text there is a signification (objects that infer in knowledge), which can be located by natural operations, such as reading, but which cannot be conveniently represented artificially.

The essential criterion of modernity in information processing methods is not the use of sophisticated and modern equipment, but the adoption of a new way of presenting the problem. The analysis of a document assumes that it is handled properly by a documentary system. This operation consists in indexing the document, describing its information content, respecting the limitations imposed by documentary

languages. Indexing is a lexical translation of language units, or even a syntactic translation when it expresses the syntactic connections among the different parts of the discourse, those that describe the content, the descriptors.

According to the use that can be made of words, they can be of four different types:

- Language Words
- Lexicon Words
- Discourse Words, and
- Words in Terminology (Terms)

Indexing is an operation that aims to represent the objects in the document. In linguistics, these objects are named referents. Thus, the words of the language or dictionary (morphemes) refer solely to their meanings, and therefore they cannot be said to point to referents.

The lexicon words, list of words of a documentary system, do not have the status of words of the language, nor of discourse words. The lexicon words do not designate objects, they refer to an open set of objects that have common characteristics. It is impossible to precisely delimit the boundaries of this set, and therefore designate a referent, to establish a direct connection between the lexicon words and an object of reality (extra-linguistic reality or imaginary reality). The subtlety that separates words from the lexicon of words into terminology explains and justifies the confusion these two terms cause.

Words in terminology introduce the frontier question. As the linguist Michel Le Guern (1989) explains, both in the lexicon and in terminology there are "words", but they are not the same words. The word "object" in the lexicon is a distinct reality, the lexicon treats words independently of things, while in terminology, words are linked to things. Words in lexicography designated as nouns are in fact predicates. They speak of qualities and not of substances. The lexicon words refer to properties, not substances, qualities, not objects. The existence of an object demands a term (word in terminology) to enter the discourse.

It is only by application in discourse and syntax that the lexicon words can take part in the construction of terms, i.e. to enter into direct relation with things (objects, referents).

This is a logical operation, and in logic the lexicon words have no extension, they have only intention (understanding). Discourse corresponds to an extensional

logic and to the words of the language an intentional logic, and one opposes the other. Extensional logic constructs segments that are bearers of reference to a reality, which designates objects of the universe, therefore classes of objects, one must consider a universe constituted of objects. Intentional logic, from language words, applies to notions (predicates) that refer to concepts.

The fact that lexicon words offer free predicates and in terminology bounded predicates, it is implied that, by combinatorial logic, it is possible to have complex predicates of both types. Also by logical operations, it is enough that one of the predicates is bound so that the complex predicate is also it.

The situation of describing the content of the documents with key words of the lexicon thus seems to be very unsustainable, even assuming the existence of a supposedly universal code (documentary language) to ensure the lexicon-discourse transition. It is necessary to seek in the discourse the necessary elements for thematic representation of discourse.

With this information, which clarifies the nature of the sign, representative of the extra-linguistic reality references, i.e., of the discourse objects, it is understood that the quality of the retrieval systems is closely related to the linking properties between discourse units and objects. This is the essence of documentary search engines. In summary, the efficiency of information retrieval systems is closely related to identifying the things that documents deal with. Documentary languages must, therefore, assume this function of pointing to the objects with pertinence, and situate the reader in this contextual universe.

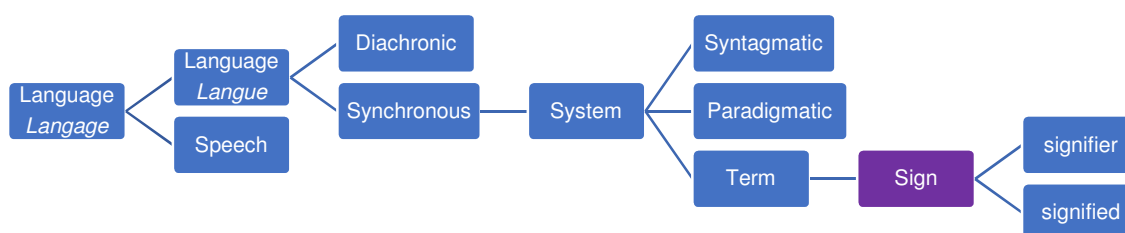
In this way, the questions raised above, regarding the use of images in catalogs are resumed, which, in turn, demand a new documentary language of the same nature, i.e. based on images.

The theoretical support structure

The two Semiology sciences, developed in Europe by Ferdinand Saussure, French linguist, and Semiotics, developed in the United States by the American philosopher Charles Sanders Peirce, both study the signs in social life, although they are different schools. Semiotics provides concepts and ideas that allow a systematic analysis of the signs, which can be words, images, complete texts, artifacts, gestures,

physical postures, etc., also studies the medium in which signs are used and how meanings are constructed.

Figure 1 - Sign and language according to Saussure: schematic representation



Parallel and at the same time (1900), indexicality theories of contents of/by images also appear in semiotics. The review of this approach converges to Charles Sanders Peirce's Theory of Signs, who defends semiotic based on the relation between the symbol, the icon and the index. Language and image are thus composed of signs that have an iconic, indexal and symbolic function. For greater objectivity of the subject treated here, the level of simplicity adopted for Peirce's theory will be centered on the triad index, icon and symbol, as follows:

- The icon is a sign in which the signifier corresponds to the meaning, for example, the image of a pen in a box indicates that inside there are pens of the same type. The image is a faithful representation of reality.

- The index is a sign represented by the resultant that meaning causes in the signifier. Thus, people with umbrellas on the street is a sign that it is raining. The representation of the rain that is not seen in the image passes through the expression of the icon(s).

- A symbolic sign exists when the relationship between signifier and signified is purely conventional, such as traffic signs, such as the dove representing peace.

It should be noted that Peirce's semiotic theory is widely applied to the description of images (BOULANGER, 1996). In image collections and archival practices, it is a literal description process of a visual document for retrieval purposes.

It is seen, therefore, that content representation through images presupposes a differentiated indexing operation. The announced advantages are multiple: user is

visual attracted by the system, intuitive understanding of the indexing code, expansion of the indexer portability, interoperability across the laws governing the discourse and indexing professional universes and, consequently, also with the user's understanding universe.

This study proposes that the methodological bases for this concept of "online library image catalog" follow the current practices and characteristics of libraries. The adoption of international classification systems is assumed: the Dewey Decimal Classification (DDC) and the Universal Decimal Classification (UDC).

For an exercise, UDC was adopted as an example. UDC is a classification based on 10 classes subdivided into 10 sub-classes each, and so on, according to the specificity of each area. The table below presents, for experimental and illustrative purposes, the representation of the UDC classes.

Chart 2 - Graphic representation of the 10 UDC classes

0	1	2	3	5	6	7	8	9
GENERALIDADES	FILOSOFIA / PSICOLOGIA	RELIGIÃO / TEOLOGIA	CIÊNCIAS SOCIAIS	MATEMÁTICA / CIÊNCIAS	CIÊNCIAS APLICADAS	ARTES / ENTRETENIMENTO	LINGUÍSTICA / LITERATURA	HISTÓRIA / GEOGRAFIA
SCIENCE AND KNOWLEDGE. ORGANIZATION. COMPUTER SCIENCE. INFORMATION. DOCUMENTATION. LIBRARIANSHIP. INSTITUTIONS. PUBLICATIONS	PHILOSOPHY. PSYCHOLOGY	RELIGION. THEOLOGY	SOCIAL SCIENCES	MATHEMATICS. NATURAL SCIENCES	APPLIED SCIENCES. MEDICINE. TECHNOLOGY	THE ARTS. RECREATION. ENTERTAINMENT. SPORT	LANGUAGE. LINGUISTICS. LITERATURE	GEOGRAPHY. BIOGRAPHY. HISTORY
01 - Bibliography and bibliographies. Catalogues	11 - Metaphysics	21 - Prehistoric religions. Religions of early societies	31 - Statistics as a science. Statistical theory	51 - Mathematics	61 - Medical sciences	71 - Physical planning. Regional, town and country planning. Landscapes, parks, gardens	81 - Linguistics and languages	91 - Geography. Exploration of the Earth and of individual countries. Travel. Regional geography
02 - Librarianship		22 - Religions originating in the Far East	32 - Politics	52 - Astronomy. Astrophysics. Space research. Geodesy	62 - Engineering. Technology in general	72 - Architecture	82 - Literature	92 - Biographical studies. Genealogy. Heraldry. Flags

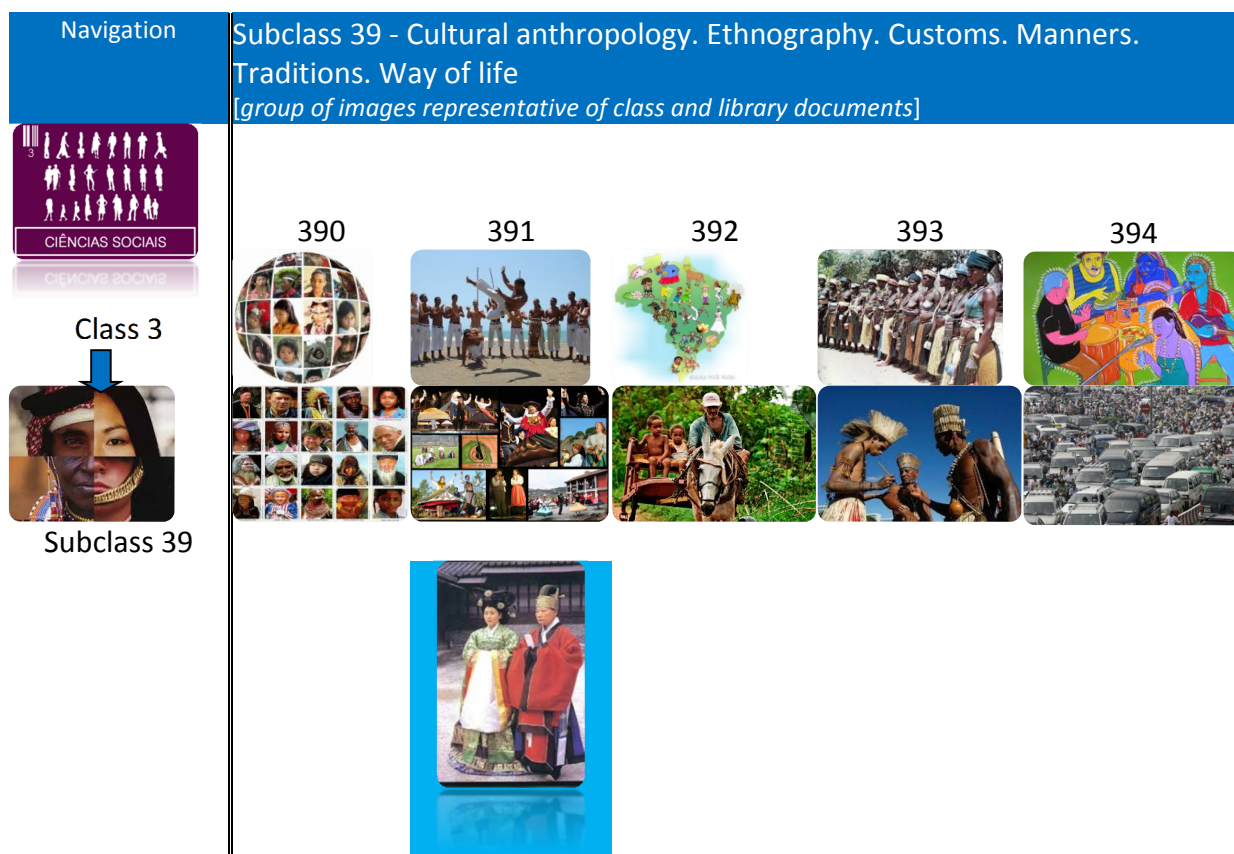
Source: adapted from public library. Azores Government.

In this graphic interface, the user's search would be given by the hierarchical (faceted) choice of images, it proposes chain navigation through the catalog subclasses (UDC) until the desired documents are obtained.

Therefore, if the user searches for a document on "Costume. Clothing. National dress. Fashion. Adornment (UDC subclass 391), the procedure would be as follows:

- The user traverses the catalog page in images until he/she finds the subject represented in the column of class 3 (Social Sciences), subclass 39.

Chart 5 - Graphic representation of UDC subclass 39



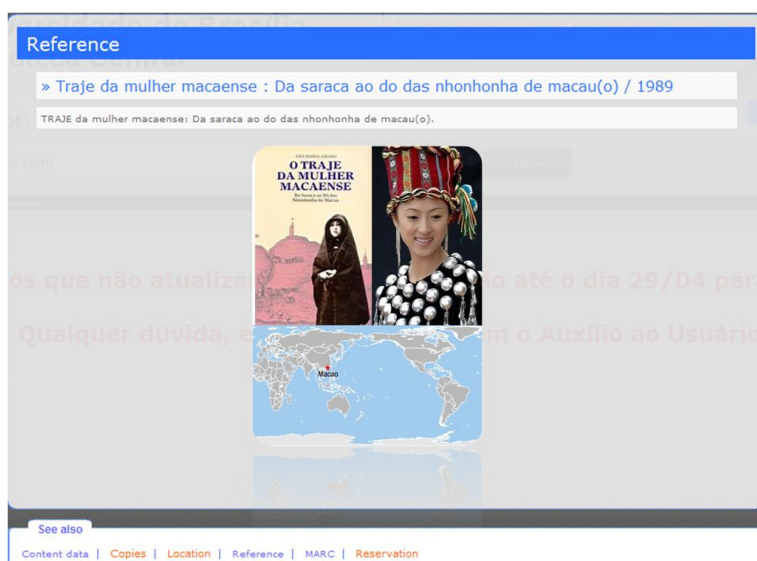
Source: by the authors (images from public domain)

Chart 6 - Graphic representation of UDC subclass 391



Source: by the authors (images from public domain)

Figure 7 - Search result contains the reference of the work in the system and its image (graphic letter)



Source: by the authors (images from public domain), from the result screen of Pergamum software

In this procedure, in which the user follows the hierarchical logic of the visual catalog or performs a random search for images, the final result must be the document reference, or a set of similar documents in images and in text, which will give access to the physical location of the documents in the collection.

This simplistic navigation representation of the visual catalog presupposes an underlying theoretical argument. According to Peirce's pragmatic theory, every object in the universe relates to an undetermined number of other objects, both directly and indirectly, in such a way that each element carries indeterminate indexical potential (LEFEBVRE, 2007, p. 220). How does the sign represent its object then, and how does it reveal its object? The iconic representation appears as a quality that possesses the object, if these qualities emerge, then the object can be identified as such and through its icon. The indexical character is closely related to the object: cold or low temperatures are closely linked to images of snow or ice. Symbolically, the sign needs to be interpreted. It is through the action of signs that a reality is inferred and in this context one can participate in indexing images. Through knowledge, the sign can be interpreted, widening the limits of the sign-object relation.

Considering the image reading task, the Saussurian and Peircian theories would be enough so that, with proper practice, an indexing professional could index images by assigning keywords according to their perception levels: signifier and signified, connotation and denotation, icon, index, or symbol. The resulting vector of these indicators would serve to store and retrieve the image document properly. But the innovation that one wants to add to the process is precisely to replace the role of the traditional keyword sign, by an image sign, which would represent the content of the documents. On this, the contribution of Jacques Bertin (1967 apud DANTIER, 2008; BERTIN, 1970) is particularly enriching.

On the analysis of information, Bertin clarifies that all thought is expressed by a system of signs, imitating a natural codification. The verbal language is a code of sound signs, the writing of a language is another type of code, and the graphic representation as well. Since the graphic representation is the transcription of the information into a graphic system of signs, the graphic representation is therefore part of the semiology.

Transcribing means separating the elements of thought from whatever system of signs is available. Regardless of the quality of this transcription, only the quality and effectiveness of graphic transcription is of interest, since it is the representation of information content. Each system of signs has its own means, style and aesthetics to represent thought. It is a relation between diverse recognized and isolated concepts, chosen from all conceivable concepts. Thus, Bertin says (apud DANTIER, 2008, page 19, emphasis added) that

Dans la représentation graphique on appellera INFORMATION le contenu traductible d'une pensée. Il est constitué essentiellement par une ou plusieurs CORRESPONDANCES ORIGINALES entre un ensemble fini de concepts de variation et un invariant.

Bertin's Image Theory also introduces the concept of image efficacy defined in relation to the perception time of the message. In order to obtain a correct and complete answer to a given question, the "construction" that requires the shortest time to perceive the response is the most effective construction, and corresponds to the image. This notion of "mental cost" was highlighted by Zipf's law (1949), widely used in library science.

In his Image Theory, Bertin evidences three successive operations for image reading that will later guide the inverse process of image construction:

1. External Identification - concretizing the proposed concepts, the components, in thought.
2. Internal identification - through which variables the components express themselves.
3. Perception of original correspondence. This perception is the result of an issue, conscious or not. The question that can be put before an information.

This means that the eye can isolate visual information by recognizing the defined correspondences and perceiving the message in its context. This perception exercise leads to the definition of image as being "the significant visual form perceptible at the minimum instant of vision".

Finally, the three basic functions of registering, communicating and processing information are retained as justifications for composing a graphic representation.

The indexing process through images thus provides a moment of reading and another of conception. Indexing through images consists of constructing an image that represents the content of the document.

- Tools equivalent to documentary languages:
 - An image bank should be constituted, preferably standardized according to an established visual programming policy.
 - The same references of the traditional instruments (Generic image, specific image, Use, Used for, View etc.) should be provided for the images, establishing standards for image use (GHEORGHITA, 2011).

- The principles announced by Bertin (apud DANTIER, 2008) must be observed to give the image corpus an identity and a language style.
 - An image bank can, such as a thesaurus, be constructed by trained professionals, in order to obtain a homogeneous and coherent collection with the area of expertise.
 - Images from this image bank will be associated metadata to facilitate the image retrieval and the transmission of their thematic characteristics to collages (composite images);
- The indexing result :
 - Contrary to keywords, the product will not be a selection of juxtaposed images, but a collage of significant images, in order to generate a "graphic letter" of the work. This graphic letter will be the symbolic sign of the work, inheriting all the characteristics of the images that compose it.
 - The composition of this "graphic letter" will be given by the minimal composition of an iconic sign and an indexical sign, closing the Peircian triad, since the "letter" itself is meant to be a symbolic sign.
 - The choice of the iconic sign should be, as much as possible, the cover image of the work, or the one that best represents it (some books do not have images on the cover).
 - The composition of the graphic letter of the work continues with the collage of the indexical and symbolic signs, accessories that allow to detail the content of the work.
 - Because of the reduced dimensions, the graphic letter is composed of few images, priority should be given to geographic and temporal images, placing the work on these two logical axes.

The image indexing experiment presented below is merely illustrative of what can be obtained as results. Because a bank of standardized images was not constructed, equivalent to the traditional documentary language, we decided to use public domain images retrieved.

Three real works, retrieved in the University Library (UL) catalog within the UDC class (39) previously exemplified, were chosen randomly to show the coherence mechanism in the indexing-retrieval process.

The information in the catalog is authored by UL.

Example 1: BOEHN, Max Von. La moda: historia del traje en europa.1945.

PRE-MARC Cataloging

Call number

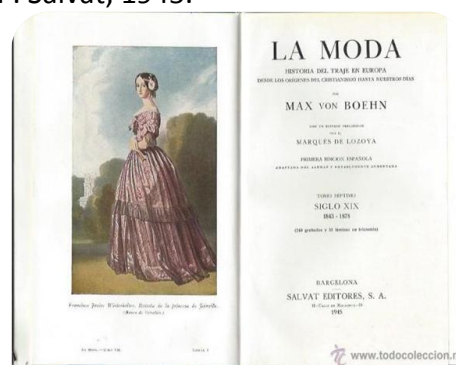
391(4)(09) B671m

Main Author

[Boehn, Max Von](#)

Main Title

Moda : Historia del traje en Europa(la)
Barcelona : Salvat, 1945.



Publication



This is the indexing card of the work, following the procedure described above: (top to bottom and left to right) Iconic image of the work, geographic index image, and two accessory images to situate the period. The result is the set called "indexing letter". Below are the indexers used by BU to describe the work.

Subject

[Adornos:antropologia cultural](#)

[Aspecto historico de determinado assunto](#)

[Elegancia:antropologia cultural](#)

[Europa](#)

[Fontes historicas](#)

[Moda:antropologia cultural](#)

[Trajes:antropologia cultural](#)

[Vestuario:antropologia cultural](#)

Example 2: AMARO, Ana Maria. O Traje da mulher macaense : da saraça ao dó das nhonhonha de macau. 1989.

PRE-MARC Cataloging

Call number

391.2(512.318) A485t

Main Author

[Amaro, Ana Maria](#)

Main title

Traje da mulher macaense : da saraca ao do das nhonhonha de macau(o)

Publication

Macau : Instituto Cultural de Macau, 1989.
199 p. : il.

Physical Description



In this case, the letter of the work was composed with only 3 auxiliary images, due to the high specialty of the theme, however completing the triad icon, index and symbol

Subjects

[Condicoes da mulher:antropologia cultural](#)

[Feminismo:antropologia cultural](#)

[Macao:china:asia](#)

[Moda feminina:antropologia cultural](#)

[Mulher:antropologia cultural](#)

[Trajes femininos:antropologia cultural](#)

[Vestuario feminino:antropologia cultural](#)

Exemple 3 - CARRILLO Y GARRIEL, Abelardo. El traje en la nueva españa. 1959.

PRE-MARC Cataloging

Call number

391(72) C317t

Main Author

[Carrillo Y Garriel, Abelardo](#)

Main Title

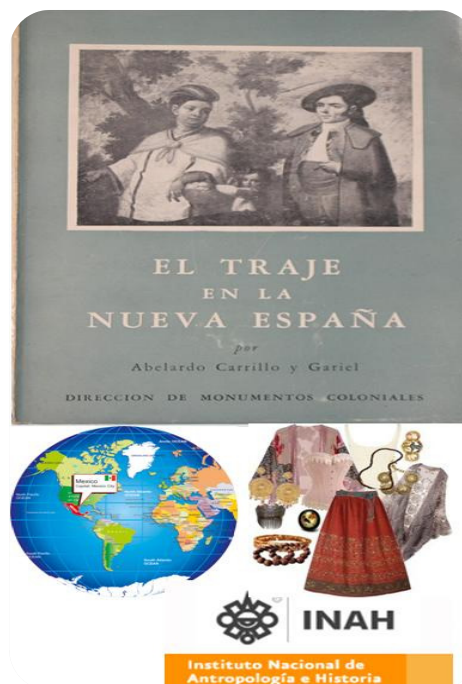
Traje en la Nueva Espana(el)

Publication

México, DF : Inst Nac Antropol E Hist, 1959.

Physical Description

207 p. : il.



Here we apply the same principles, with the introduction of a fourth image that places the user in the dimension of authority of the work: the Institute of anthropology and history. We can also observe the translation of the indexing element when one reads in the work "Nueva España" that today is Mexico.

Subjects

[Adornos: antropología cultural](#)

[Elegancia: antropología cultural](#)

[Mexico: america do norte](#)

[Moda: antropología cultural](#)

[Trajes: antropología cultural](#)

[Vestuario: antropología cultural](#)

Conclusions

Content representation through images is supported by theories that base both Saussurian semiology and Peircian semiotics. Image description as their textual representations show the bases for management systems of image databases, essential tools for the composition of the principle of image indexing announced in this work. The revisited guidelines of Bertin show how one can get to an image composition perfectly in agreement with the linguistic theories that clearly explain the process of textual indexing and therefore the mechanisms of the indexing through image.

The proposed examples for navigation in image OPACs, as well as the composition of the descriptive records of documents, confirm the application of the cited theoretical bases and clarify the feasibility of a more effective indexing. In these

terms, the results open up multiple possibilities for accessibility to documentary collections. The target audience gets broader, and the greater the opportunities for developing reading habits. Image collections such as the publishing industry of image, and in particular the comic books, discover new forms of expansion in this proposal. Information professionals develop new skills with direct repercussions on the internal structure of information management systems, and on information retrieval as well. In the same way, web search systems will be enriched by these indexing and information retrieval strategies.

The proposed methodology thus opens the field for new interface developments that are more adapted to communities with special needs, the disabled and the functional illiterates. Further studies on the mechanisms of semantic perception of images are expected. Transdisciplinary competences in neuroscience, linguistics and information sciences may together explore such issues as the nature of the cognitive processes of the deaf in information organization and classification, the functioning of autism communication grammars, the mechanisms of action of bibliotherapy for the elderly, patients with Alzheimer's, other depressive disorders, among others. Finally, the use of image to establish links between users and documents seems to find in information retrieval systems a vast field of scientific verification both in research and in development.

References

The references was made following the ABNT rules.

ABNT. *Métodos para análise de documentos: determinação de seus assuntos e seleção de termos de indexação*. Rio de Janeiro: ABNT, 1992. 4p.

ASHTON, Danny. *10 reasons why you should care about visual content marketing*. Blog. Publicado em 19 de fevereiro, 2015. Disponível em: <<http://neomam.com/blog/13reasons/>>. Acesso em: 5 maio 2015.

BARTHES, R. *Le système de la mode*. Paris: Seuil, 1967.

BERTIN, Jacques. La graphique. *Communications*, v. 15, n. 1, p. 169–185, 1970.

BIBLIOTECA pública. Governo dos Açores. Disponível em: <<http://www.bparah.azores.gov.pt/html/bparah-seccao+infanto-juvenil+cdu.html>>. Acesso em: 3 maio 2015.

BOULANGER, Chantale. La surimage. *Applied Semiotics/Sémiotique appliquée*, v. 1, n. 2, p. 159–172, 1996.

BRITO, Marcilio de. Sistemas de informação em linguagem natural: em busca de uma indexação automática. *Ciência da Informação*, v. 21, n. 3, p. 223–232, 1992.

DANTIER, Bernard. La représentation et l'étude visuelles des informations. In.: BERTIN, Jacques. *Sémiologie graphique, les diagrammes, les réseaux, les cartes*. 4.ed. Paris. Ehess, 2008. 32 p.

GHEORGHITA, Inga. Methodologie de construction automatique du thesaurus pour l'indexation et la recherche des images. *RECITAL*, n. 2, p. 221–228, Jun. 2011. Disponível em: <<https://hal.archives-ouvertes.fr/hal-00695722>>. Acesso em: 2 maio 2015.

GREIMAS, A.J. *La mode en 1830: Essai de description du vocabulaire vestimentaire d'après les journaux de mode de l'époque*. Paris, Sorbonne (thèse dactylographiée). s.d.

GREIMAS, A.J. *Sémantique Structurale recherche de méthode*. Paris: Larousse, Langue et Langage, 1966.

GREIMAS, A.J. *Du Sens. Essais Sémiotiques*. Paris: Seuil, 1970.

GREIMAS, A.J. Les Actants, les Acteurs et les Figures, In: *Sémiotique narrative et textuelle*, Claude Chabrol (dir.), Paris, Larousse, 1973, p. 161-176.

GREIMAS, A.J. *Sémiotique et Sciences Sociales*, Paris, Seuil, 1976. 219 p.

GUIRAUD, P. *De la grive au maquereau : le champ morpho-syntaxique des noms de l'animal tacheté*. Le français moderne. v.34, 1966

LE GUERN, M. *Sur les relations entre terminologie e lexique*. META Journal de traducteurs. Montréal. 1989

LEFEBVRE, Martin. The art of pointing: on Peirce, indexicality, and photographic images. *Photography Theory*, p. 220-244, 2007

MOUNIN, G. Essai sur la structuration du lexique de l'habitation. *Cahiers de lexicologie*, v.5, 1965.

MOUNIN, G. Un champ sémantique: la dénotation des animaux domestiques. *La linguistique*. v.1, 1965.

ROGERS. Simon. *What fuels a tweet's engagement? (@smfrogers)*. Data editor 03/10/2014. Disponível em <https://blog.twitter.com/2014/what-fuels-a-tweets-engagement> Acesso em abril, 2015.

SOCIAL BAKERS. *Photos cluttering your facebook feed? here's why pictures account for majority of brand facebook posts*. Abril 21, 2014, Advertising & Marketing. Disponível em <http://www.emarketer.com/Article/Photos-Cluttering-Your-Facebook-Feed-Here's-Why/1010777/1> Acesso em maio, 2015.

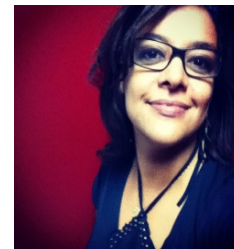
UNISIST. Princípios de indexação. *Revista da Escola de Biblioteconomia da UFMG*, Belo Horizonte, v. 10, n. 1: p. 83-94, mar. 1981.

ZIPF, George K. *Human behavior and the principle of least effort*. 1949.

Interactive conceptual maps as a didactic instrument in Information and Knowledge Organization



Marilda Lopes Ginez de Lara
University of São Paulo
larama@usp.br



Gabriela Previdello
University of São Paulo
In memoriam



Nair Yumiko Kobashi
University of São Paulo
nairkobashi@gmail.com

Introduction

Alongside the conceptions that underpin the enterprise of organizing information and knowledge to make them accessible to users, it is also essential to invest in training people who perform informational activities in a consistent way, in tune with the demands of different social segments. It is in this spirit that we develop the project presented here, in the belief that training should be emancipatory, replacing the pattern based on the linear and unidirectional transmission of knowledge, sometimes practiced in the area. We believe that the current role of the professor is, first of all, of mentoring and guidance, providing the learner with a set of knowledge in order to creatively carry out the learning experience.

Adopting these principles and availing ourselves of the possibility of accessing the technological resources offered by São Paulo University of, within the Pró-Ensino Program, we developed the project "Information and Knowledge Organization: interactive conceptual maps in the construction of documentary information". The project seeks to integrate the disciplines of Information and Knowledge Organization

from the LIS Department of the School of Communications and Arts of São Paulo University - ECA / USP, based on shared principles and procedures. With this activity, we intend to deepen the discussion regarding alternative modes of academic and professional training in the area, providing didactic resources in the information and communication network environment.

The field of Information and Knowledge Organization has, among its objectives, to reflect critically and to develop the theoretical and methodological bases of the organization, location, retrieval and use of registered knowledge (CINTRA et al., 2002; HJORLAND, 2018). The denomination of the field, as well as its organization in disciplines, varies greatly among the undergraduate courses in Librarianship due to the theoretical-methodological aspects chosen by each institution. If the variation can be healthy, the formation of the student is problematic, as the student does not always understand the field as an integrated and interrelated set of processes. Segmentation in different disciplines therefore requires an integrative oriented work.

In this perspective, the project seeks to approach teaching-learning in an open way, systematizing contents that relate to theories, methods and construction processes of documentary information, accessible on digital platforms. Its justification is the need to gather, for the student (and the professor), a set of references of different types and levels of presentation to complement the studies carried out in person (regular learning). It uses, for this purpose, instruments of the field of Information and Knowledge Organization, allowing the access to multiple concepts and the identification of their relations.

Objectives and rationale

The general objective of the Project "Information and Knowledge Organization: interactive conceptual maps in the construction of documentary information" is, initially, to offer knowledge, in a digital environment, to undergraduate students of the Library and Information Science Course of ECA/USP. It includes the contents of the following disciplines: Introduction to Documentary Analysis, Elements of Logic for Documentation, Documentary Languages I and II, Documentary Linguistics, Indexing: theory and methods. The disciplines, offered between the 1st. and the 5th. semesters of the course, taught by the teachers participating in the project.

Designed to be developed in phases, the project must be gradually enriched with new contents and resources, in addition to being subjected to successive evaluations.

The specific objectives of the project are:

- to develop learning resources related to the main topics of the disciplines, emphasizing their concepts, work methodologies, workflows, highlighting the relationships across the documentary operations;
- to introduce the concepts of Documentary Analysis focusing on the processes of text condensation and representation (verbal and visual);
- to introduce the concepts of Logic, Linguistics and Terminology by exploring the differences between Natural Language and Documentary Language;
- to present the main indexing operations of textual and image documents, exploring documentary representation processes through Documentary Languages;
- to discuss the structural aspects of Documentary Languages and vocabulary control methodologies;
- to introduce the principles of promoting the interoperability of vocabularies for their sharing and reuse in the Semantic Web environment;
- to present examples and suggest activities to apply the methodologies.

Theoretical-methodological principles

The assumption of the project is that the activities of Information and Knowledge Organization necessarily develop in the universe of language, which is why their theoretical references are based, among others, on the Sciences of Language, Logic, and on Terminology. The field of Information and Knowledge Organization aims to organize the knowledge objects in their singularity (texts, image, statistical tables, etc.), kept in systems, to make them available for access and use in various media (files, catalogs, databases, websites), observing the different levels of data granularity (individual objects, complex objects organized in systems etc.) as well as the user's cultural and language references. Likewise, it aims to reflect on such practices to understand its fundamentals, commitments and ethical positioning.

In the context of language studies, the concepts of Natural Language and Documentary Language (or artificially constructed language), the notion of linguistic structure (minimum structure of signification), of structuring (syntax) and of use

(pragmatic), among others, are fundamental. Also essential are the knowledge about the nature of language, its symbolic function, the phenomena of enunciation, discursiveness and textualization, as well as the distinction between the expansion and condensation operations and the logical-semantic organization of vocabularies to represent informational contents. From Semiotics, concepts of object, sign, semiosis, interpretant, context and interpreter are important, and from Terminology, concepts of knowledge domain or area of activity, formation and exploration of corpora, term, concept, relations among concepts, domain tree, relations of conceptual and linguistic equivalence and of social validation of terms. From Logic, the notion of concept, its logical ordering, is essential, based on the identification of extension and intension of the concept properties. The notions of conceptual equivalence and space and time relationships to operationalize and enrich the organization of descriptors in information representation tools for retrieval are also important.

Finally, the field of Information and Knowledge Organization seeks to take care of the aspects related to signification and relations of meaning, recognizing that the characterization of documentary discourse differs from discourses in natural or specialized language. By seeking to represent the texts for retrieval purposes, the field uses operations that ensure relations of meaning through oppositions and contrasts, hyponymy or subordination, hyperonymy or superordination, generic and collateral displacements, part-whole relations. Via Terminology, the field of Information and Knowledge Organization updates the idea of semantic field by circumscribing its vocabularies to special domains and providing them with structuring from pragmatic references.

The interdisciplinary attitude seeks to guide the organization of the field: concepts are not merely borrowed from other domains; rather, they are evoked to better understand the linguistic-communicational manifestations as well as to underpin the methodological operations of the specific domain of Information and Knowledge Organization.

From a didactic-pedagogical point of view, we assume that knowledge construction requires articulating the conceptual to the empirical. In fact, the field is not restricted to the observation and interpretation of phenomena. The commitment is to the creation of theories and methods to treat information for social uses. Information and knowledge organization finds its field of validation, therefore, in the effectiveness with which information can be retrieved and used socially.

Recent discussions on the Semantic Web have revitalized the basic theories of concept organization. Interoperability investments made by W3Consortium, IFLA and consolidated in international standards of Documentation, such as ISO standards (ISO 25964, 2011; 2012), have emphasized the importance of concept and conceptual systems as a basis for representation and organization of controlled vocabularies.

Method

The concepts of Knowledge Organization and Representation presented in this phase of the project were selected according to their importance to the subjects of the Librarianship Course at ECA-USP. The syllabus' contents and its bibliography were the corpus of identification and selection of concepts. For the relations among the concepts to have greater visibility, a fact not always perceived by the students, they were organized in thematic categories and not by disciplines. We sought to understand such categories in a pragmatic sense: groupings that manifest particularities in relation to an area of knowledge (in our case, Information and Knowledge Organization) that serve institutional purposes or particular points of view, and are not, therefore, of universal application.

Operationally, the selected concepts were, in a first moment, organized as a thesaurus, since the construction methodology of these languages proved to be adequate to house and distribute the terms structurally. In order to store the terms that correspond to concepts, we first resorted to Multites (Multisystems, 2005-2014) and then to TemaTres (TemaTres, 2015), considering not only the fact that the latter is open and free software, with the advantage of being designed to promote interoperability across different controlled vocabularies.

The project is structured in modules that present the following basic contents:

- 5-10 minute video lessons;
- specific literature: files with fundamental texts;
- references, sources of information;
- examples of application;
- addresses of websites with related content;
- related software;
- exercises.

The contents are structured in the form of a conceptual map, with the difference that their design is based on hierarchy, association and equivalence to ensure greater

consistency to the relationship network, according to the methodology for organizing a thesaurus. The main nodes of the map correspond to thematic categories which, in turn, understand concepts and their relationships with other concepts within the category itself, or with concepts from other categories. The central concepts are, in turn, linked to bibliographical references, examples, exercises, etc.

The used technology support includes a content management system for CMS website in Wordpress, with layout in HTML4.0 Transitional / CSS / PHP 5.2 / JS languages, and MySQL 5.1 database, in addition to other development standards such as JQuery version 1.5x and Javascript version 1.1x. The included technological requirements were usability, portability and interoperability, so that the languages were widely understood and compatible with all browsers. The system operates in an integrated way to the tools in use at São Paulo University, such as the Moodle platform - STOA - USP, and specialized software related to disciplines (development of controlled vocabularies, conceptual maps, information visualization) (YouTube, Vimeo, Soundcloud) and social media (Facebook, Google+, Twitter).

The graphic design project sought to equate the teaching/learning proposal using alternative forms of visualization of contents by grouping, in the same space, the concepts of the various disciplines in thematic categories using solid colors to highlight them. The implementation of the project started with 200 theoretical and operational terms selected from the programmatic contents and basic bibliography of the disciplines. The terms were then organized into a standardized structure, which integrates, for each term, in addition to their network of relationships, the conceptual definition or characteristics of the concept, term families, use examples, exercises, videos, etc. The adopted structuring form allows the student, when choosing a term or a thematic category, to discover the multiple relationships that exist between the theoretical and operational concepts, as well as the corresponding information that contextualizes them.

The information is distributed in a panel that allows the updating of the data in a practical and intuitive way. The public reading of the data can be done through an API (<http://www2.eca.usp.br/oc/sobre-api/>), and, in the near future, the information will be made available to other institutions and users.

For the following phases, spaces are provided for virtual conferences and discussions between students and teachers, as well as for hosting criticisms, observations, questions and suggestions. Eventually, the modules should also contain

films, videos with reports of experience from other institutions, proposals for solutions for comparison, virtual tours.

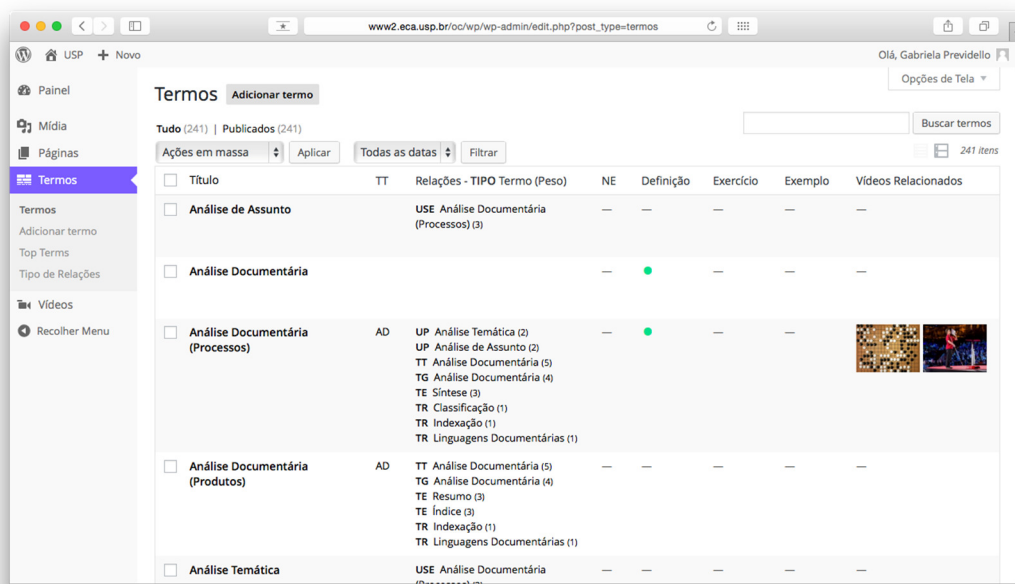
Results

We present, in this section, the working platform and the result of the preliminary organization of data and partial visualizations of categories and terms in the form of conceptual maps.

Work platform

The work platform has the following functionalities:

Figure 1 - Work screen: inclusion of terms and their relations



Organization of terms

The contents of the preliminary data organization can be visualized as follows:

Figure 2 - Alphabetic visualization of terms

MAPA CONCEITUAL INTERATIVO	A	Classificação	Disjunção	Falácias Dedutivas	Intuição Empírica	Nome	Proposição Negativa	Relação de Instanciação	Sinonímia	Substância da Expressão/Substância do Conteúdo
LISTA ALFABÉTICA DE TERMOS	Análise de Assunto	Classificação Bibliográfica de Bliss	Documentação	Falácias Indutivas	Intuição Intelectual	Nota de Escopo	Proposição Particular	Relação de Interseção	Sinónimo	Substância do Conteúdo
SISTEMAS DE ORGANIZAÇÃO DE TERMOS	Análise Documentária	Classificação da Biblioteca do Congresso	E	Forma da Expressão	Intuição Sensível	Notação	Proposição Possível	Relação de Negação	Síntese	
LINGÜÍSTICA	Análise Documentária (Processos)	Classificação de Assuntos Brown	Eixo Paradigmático	Forma da Expressão/Forma do Conteúdo	L	O	Proposição Predicativa	Relação de Oposição	Sistema de Classificação Bibliográfica de Bliss	T
TERMINOLOGIA	Análise Documentária (Produtos)	Classificação Decimal de Dewey	Eixo Sintagmático	Forma do Conteúdo	LCSH	Ontologias	Proposição Singular	Relação Funcional	Sistema de Classificação da Biblioteca do Congresso	Taxonomias
ANÁLISE DOCUMENTÁRIA	Análise Temática	Classificação Decimal Universal	Enunciados Lógicos	Generalização Estatística	Língua/fala	Operadores Booleanos	Proposição Subalterna	Relação Genérica	Sistema de Classificação de Assuntos Brown	Terminologia
LÓGICA	Analogia	Classificação Documentária	Equação de Busca	Generalização Universal	Linguagem	Palavra-Chave	Proposição Universal	Relação Gênero/especie	Sistema de Classificação de Assuntos Brown	Termo
SÍMBOLOGIAS E TIPOS DE RELAÇÃO	Argumento Causal	Classificação dos Dois Pontos	Estrutura de Busca	Glossário	Linguagem Artificial	Paradigma/síntagma	Proposição Universal Afirmativa	Relação Hierárquica	Sistema de Classificação Decimal de Dewey - CDU	Termo Associado
VÍDEOS	Argumentos	Classificação Expansiva de Cutte	Especificidade	Generalização Universal	Linguagem de Especialidade	Pensamento Lógico Ver Raciocínio	Proposição Universal Negativa	Relação Lógica	Sistema de Classificação Decimal de Dewey - CDU	Termo Específico
EXERCÍCIOS	Argumentos Contra o Homem	Conceito	Estrutura Linguística	Hipótese	Linguagens Documentárias	Poder Explicativo	Proposições Lógicas	Relação Ontológica	Sistema de Classificação Decimal de Dewey - CDU	Termo Genérico
SOBRE O PROJETO	Argumentos de Autoridade	Conceito Específico	Estrutura Mínima da Significação	Hipótese Ad Hoc	Linguística	Pós-Coordenação	Quadrado dos Opostos	Relação Paradiigmática	Sistema de Classificação dos Dois Pontos	Termo Específico Partitivo
CONTATO	Argumentos Dedutivos	Conceito Geral	Exaustividade	Hipótese Auxíliar	Linguística Aplicada	Pragmática	Quase-Sinónimo	Relação Sequencial	Sistema de Classificação dos Dois Pontos	Termo Genérico
API	Argumentos Indutivos	Conceito Individual	Explicação	Hipótese Preliminar	Linguística Descritiva	Pré-Coordenação	Raciocínio Dedutivo	Relação Sinotagmática	Sistema de Classificação dos Dois Pontos	Termo Médio
ORGANIZAÇÃO DA INFORMAÇÃO E DO CONHECIMENTO	Assunto	Conceito Partitivo	Extensão	Homonímia	Linguística Documentária	Princípio da Identidade	Raciocínio Indutivo	Relação Termo-Notião	Sistema de Classificação dos Dois Pontos ou Colon Classification	Termo Não-Preferido
MAPAS CONCEITUAIS INTERATIVOS NA CONSTRUÇÃO DE INFORMAÇÃO DOCUMENTÁRIA	Cabeçalhos de Assunto	Conjunção	Faceta	Indeção	Listas de Autoridades	Princípio da Não Contradição	Raciocínio Lógico	Relevância	Sistema de Classificação Expansiva de Cutte	Termo Relacionado
ESCOLA DE COMUNICAÇÕES E ARTES DEPARTAMENTO DE BIBLIOTECOMIA E DOCUMENTAÇÃO	Característica	Falácia Causal	Falácia da Afirmação Consequente	Indeção Automática	Listas de Cabeçalho de Assuntos	Princípio do Terceiro Excluído	Ração Objetiva	Representação de Assunto	Sistema de Classificação Expansiva de Cutte	Tesouros
	Características Accidentais	Falácia da Composição	Falácia da Divisão	Indeção Manual	Lógica Operacional	Princípios da Lógica	Ração Subjetiva	Representação de Assunto	Sistema de Classificação Expansiva de Cutter ou Cutter Classification	Top Term
	Características Complexas	Defluição	Falácia da Estatística Insuficiente	Indeção Semi-Automática	Lógica	Princípios da Razão	Resumo	Representação Temática	Sistema de Classificação Bibliográfica	Tratamento da Informação
	Características do Conceito	Definição por Extensão	Falácia da Estatística	Índice	Método Hipotético Dedutivo	Problema	Ritual Modal	Resumo	Sistemas de Classificação Bibliográfica	Validade dos Argumentos
				Indução por Enumeração	Mononímia	Proposição Afirmativa	Relação de Coordenação	Semântica	Sistemas de Conceitos	
						Proposição Contraditória	Relação de Disjunção	Significante/significado	Sistemas de	

Figure 3 - Graph visualization of the category "Knowledge organization systems"

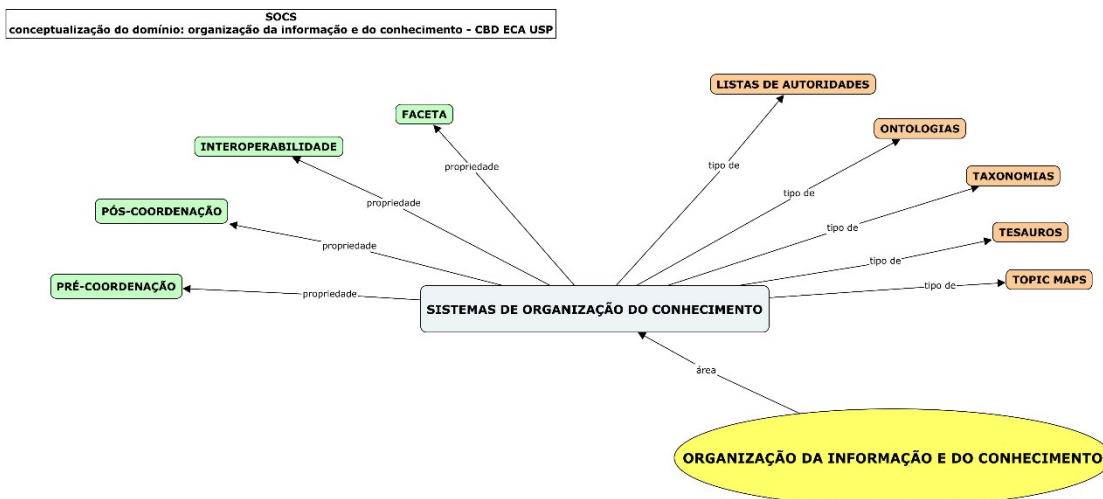
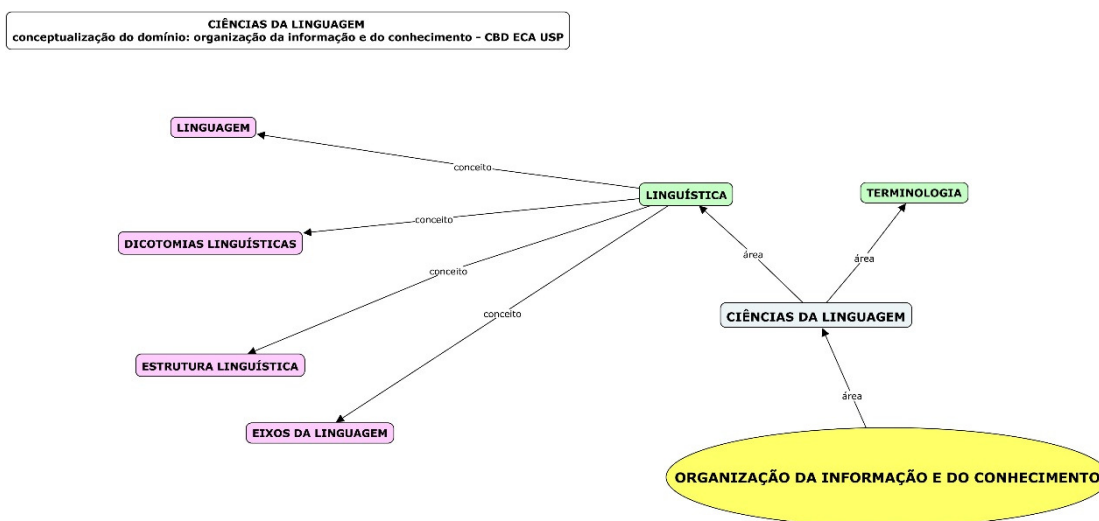


Figure 4 - Graph visualization of the category "Language sciences"



The development of the project has shown that, in addition to the benefits for undergraduate students, it triggers a process of productive discussions about the didactic-pedagogical plan among the teachers, especially in relation to the privileged contents and the integration among the disciplines. Likewise, the project highlights the need to invest in the production of a scheme of equivalences between concepts that stems from different theories that coexist in the area, as well as in the fundamentals of the choices made.

Another important aspect is the validation that must occur in 2 continuous semesters, scheduled for the 2nd. semester 2015 and 1st. Semester 2016, minimum period to ensure that the student attends at least 2 disciplines and can interact with the conceptual maps available on the platform for complementing their training.

Conclusions

The renewal of the teaching of Information and Knowledge Organization faces many challenges: the conception of a way to organize contents, with didactic purposes; the determination of nuclear and peripheral concepts and their forms of interaction; skills needed to implement the proposal; the choice of appropriate technologies for these purposes and the raising of resources for teaching activities, which are known to be more scarce than the financial contributions available for research.

The choice of the proposal axis - the categorization of concepts - was defined as it allows to show the main concepts and operations of the Organization of Information and Knowledge in an integrated, articulated way. With this approach, we tried to overcome the fragmented vision of the field. In fact, if the disciplinarization, or

reduction of knowledge in disciplines allows to deepen knowledge, it may, on the other hand, promote the dispersion of the concepts and operations of a given domain. It is necessary, therefore, in the formation, to invest in actions that promote an integrating look. The use of different graph visualizations, as exemplified above, is certainly an important differential of this proposal.

For the exploitation of the resources offered by the digital environment, several forms of content presentation were adopted: video lessons; links to key texts; information sources; examples of applications, sites with related content; related software; exercises.

The used technology support includes a content management system for websites, appropriate languages, relational databases. The technological requirements were selected based on the concepts of usability, portability and interoperability. Finally, the system operates in an integrated way with the tools in use at São Paulo University, such as the Moodle platform - STOA - USP and social networks.

We hope that the results of this project can be used by the community teaching Information Organization, which is indispensable for the training of librarians and other information professionals.

Acknowledgement

We wish to thank the current project team, especially Profs. Dr. Cibele Marques dos Santos, Dr. Giovana Deliberali Maimone, Dr. Vânia M. Alves Lima, and the trainee (undergraduate student) Lúcia Marques. We also wish to thank the Pro-Teaching Program of USP for the resources granted for the development of the project.

References

The references was made following the ABNT rules.

CINTRA, A.M.M.; TÁLAMO, M.F.G.M.; LARA, M.L.G.; KOBASHI, N.Y. *Para entender as linguagens documentárias*. São Paulo : Polis, 2002.

HJORLAND, B. Knowledge Organization Systems (KOS). In: HJORLAND, B. *Lifeboat of Knowledge Organization*. 2008. Disponível em: <http://www.iva.dk/bh/lifeboat_ko/CONCEPTS/knowledge_organization_systems.htm> Acesso em mar. 2015.

INTERNATIONAL STANDARD ORGANIZATION. *ISO 25964: thesauri and interoperability with other vocabularies. Part 1: Thesauri for information retrieval*. Geneve: International Standard Organization, 2011.

INTERNATIONAL STANDARD ORGANIZATION. *ISO 25964: thesauri and interoperability with other vocabularies. Part 2: Interoperability with other vocabularies*. Geneve: International Standard Organization, 2012.

MULTISYSTEMS, 2005-2014. Disponível em: <http://www.multites.com/>>
Acesso em jul. 2015.

TEMATRES, 2015. Disponível em: <http://www.vocabularyserver.com/index.html> Acesso em jul. 2015.

Domain analysis in Knowledge Organization: exploring thematic and citation relations



Bruno Henrique Alves
São Paulo State University
brhenriquealves@gmail.com



Ely Francina Tannuri de Oliveira
São Paulo State University
etannuri@gmail.com



Maria Cláudia Cabrini Grácio
São Paulo State University
cabrini@marilia.unesp.br

Introduction

Knowledge Organization is a discipline that addresses the development of techniques for construction, management and use, evaluation of scientific classifications, taxonomies, nomenclature and documentary languages. It also deals with natural language use and retrieval methodologies (NAVARRO, MARCO, 1995, BARITÉ, 2001). In this way, the object of study of Knowledge Organization, according to these authors, is the knowledge that, socialized, can generate new knowledge.

Currently, the study of its domain has been the target of several studies, such as: Guimarães; Santos; Cândido; Pinho (2014), Smiraglia (2014) and Arboit (2012), among others, who propose to define and conceptualize its scope, its object of study, its theoretical foundations and, especially, to outline methodologies that can best understand the knowledge constructed.

Hjørland and Albrechtsen (1995) developed an epistemological social paradigm named Domain Analysis (A.D.), whose main aspect highlights the relationship between

different fields of knowledge and discursive communities. It is a process by which one can verify what is significant in some scientific community, pointing out elements that allow analyzing their scientific context, such as trends, patterns, objects (people), existing processes and relationships, involving the cultural and anthropological, historical and social understanding.

In this research, domain is the area of knowledge named KO, in which the boundaries are demarcated, with professionals or groups articulated in thought and language.

Among the 11 domain analysis approaches presented by Hjørland (2002), we highlight bibliometric studies, which, combined with epistemological or historical studies, produce a fertile interaction and better visualization of different scientific areas, revealing dominant theoretical currents, either through scientific collaborations, or through the analysis of citations or co-citations.

The set of bibliographic references used in the elaboration of a document can be analyzed as a reflection of a discursive community composed of researchers and their thematic, in order to constitute a scientific domain.

Leydesdorff (1998) notes that citations point to both the social and cognitive context of knowledge in scientific practices, sustain scientific communication, and define the intellectual structure of an area of knowledge.

Citation analysis consists of a way of defining a domain, by identifying the authors of greater insertion and impact, the dialogue and the correlation established among them by the citing community (SMIRAGLIA, 2011; GLÄNZEL, 2003). It points out its paradigms, relevant methodological procedures and the "vanguard" researchers.

Considering the above, this research aims to analyze the researchers that contribute to the development of Knowledge Organization (KO) through domain analysis, in its approach bibliometric studies.

Specifically, it aims to identify and present, in a diachronic way, the scientific elite in KO, characterized by the most productive researchers in the journal Knowledge Organization. Also, it aims to highlight the themes contemplated by the most productive researchers and their theoretical and methodological references, as well as the network of interlocution of this elite with the most cited authors, in the 2004-2008 and 2009-2013 periods.

This research is justified by its analytical contribution to the understanding of the domain Knowledge Organization, by offering theoretical and methodological subsidies related to the scientific elite and their respective referents.

Methodology

The journal Knowledge Organization, indexed in Scopus database was used, and we retrieved 41 articles published by 57 authors in the 2004-2008 period, and in the five-year period 2009-2013, we retrieved 74 articles published by 105 authors. We selected the 19 most productive authors, corresponding to the publication of at least 2 articles, considering the period of 10 years. Subsequently, the publications of these 19 researchers were separated in the two 5-year-periods, for a more detailed diachronic analysis in relation to the citations.

In view of the keywords, the themes in the 36 articles of the 19 researchers were grouped using Dahlberg's Classification System for Knowledge Organization Literature in order to identify thematic proximities among the most productive authors.

The authors were grouped according to the distribution of the production by 5-year-period and classification of the thematic, according to Dahlberg's system, through the analysis of clusters, using Ward method and Euclidian distance. The visualization of the groups of researchers is presented in Table 1, along with the production per five-year period, total and distribution of the articles by Dahlberg's category, in a non-excluding way.

In the first five-year period (2004-2008), for the 19 most productive researchers, we found 651 citations, related to 511 cited authors; and in the second five-year period (2009-2013), 775 citations, relative to 780 cited authors.

The most cited authors were those who received at least 4 citations, totaling 18 authors in the first five years and 10 researchers in the second period. Self-citations were eliminated, and researchers who were referred to with more than one reference in the same article were recorded only once.

Finally, the two asymmetric matrices 19x18 and 19x10, related to the most productive and most cited authors were constructed using Excel software, one for each five-year period. Ucinet software was used in order to generate two mode networks (citing and cited).

Presentation and analysis of results

Table 1 presents the 19 most productive researchers, per five-year period and thematic work, grouped according to the proximity of the Cluster multivariate analysis.

Table 1. Most productive researchers in Knowledge Organization, per five years, total and according to Dahlberg's System Classification for KO Literature

Researchers	2004 a	2009 a	Total	C1	C2	C3	C4	C5	C6	C7	C8	C9
Gnoli, C. G1	3	1	4	3	1	2					1	
Kipp, M.E.I.	0	3	3	1	2	1					2	
Campbell, D.G.	0	2	2		2	1					1	
Mazzocchi, F. G2	1	1	2		2	1					1	
Osińska, V.	0	2	2		1	1	1					
Olson, H. A.	1	2	3		2	2	1				1	
Dahlberg, I.	1	1	2	2	1							
Tennis, J. T. G3	1	1	2	1		1						
Buchel, O.	0	2	2	1							1	
Mai, J.-E.	0	2	2		1						1	1
Golub, K.	1	1	2		1	2					1	
Lee, H.-L. G4	1	1	2		1	2					1	
Souza, R.	1	1	2			2					1	
López-Huertas, M.J.	2	0	2		1						1	1
Ménard, E. G5	1	1	2			1					2	
Park, J.-R.	1	1	2								2	
Chaudhry, A. S.	2	1	3		3			1				
Khoo, C. G6	1	1	2		2							
Wang, Z.	1	1	2		2							

G1: Total of outstanding articles and production in 4 different categories;
 G2: All researchers published in the 2nd quinquennium and in C2 and C3;
 G3: Presence in the 2 quinquennia and dispersed in the categories;
 G4: All researchers published in the 2 quinquennials and in C3 and C7;
 G5: All researchers with articles in C7 and little presence in the others;
 G6: All researchers with articles in the 2 quinquennia and centered in C2
 Source: by the authors

It is observed that five researchers had no production in the first five-year period, and in the second period only one researcher did not present production, meaning an expansion and greater dispersion of the producing elite in this five-year period.

In addition, out of the 19 researchers, 10 of them published an article in the first-five years and an article in the second five-year period, evidencing a relative stability of research in the area.

In the last ten years, the five researchers who entered the second period - Kipp, Buchel, Campbell, Mai and Osinska - are the ones with the highest production, along with Olson, who was already in the first five-year period group. On the other hand, the

most productive researchers in the first five years decreased their production significantly in the second five-year period.

Moreover, it is noted that the large producers in KO who published in K.O. in this period come from the U.S., Canada and Europe, highlighting one single Brazilian author - Souza - from the Getúlio Vargas Foundation.

In relation to the most addressed themes in the articles, from the classification used, the most contemplated is C2, related to the Classification Systems and Thesauri (CS & T). Structure and Construction, which involve relations among concepts, taxonomies and indexing language, among other items . Next, C3 related to the Classification & Indexing (C&I) (Meth.), which includes indexing and classification theory, subject analysis, automatic indexing and generation of indexing and programs, and also C7, Knowledge Representation by Language and Terminology, which addresses natural language in relation to KO, semantics, online retrieval systems, technologies and terminologies. On the other hand, it is observed the absence of works in C6, referring to On Special subjects CS & T, and only an article by Chaudhry in C5, On Special Objects CS (Taxonomies), in the different areas of Knowledge, and C8, Applied Classing & Indexing (C&I).

From the results, we verify a greater concern of KO researchers to turn to the object of the area, deepening the studies in order to consolidate the fundamental themes of the area, overlooking less prioritized themes applied in other areas of knowledge .

Figures 1a and 1b show the two networks of citing-cited researchers, in which the red circles are the citing ones. The thickness of the straight segments represents the citation frequencies. Note that there are five authors who appear cited in the two five-year periods, namely Svenonius, Albrechtsen, Hjørland, Mai and Olson, outstanding researchers in the study of the foundations of Knowledge Organization in the twentieth century.

Svenonius concentrates his studies in the metaconstruction for the theory of Knowledge Organization, whose seminal work is the book *Intellectual Foundations of Information Organization*. Albrechtsen focuses his studies in the following areas of research: knowledge management, domain analysis, entrepreneurship. Hjørland specializes in information architecture, information literacy, information retrieval and informational behavior. Mai studies questions about conceptual, methodological and programmatic constructions, as well as the nature of classification. Olson's research

focuses on the critical analysis of subject representations and classification systems, with a feminist, post-structural, and postcolonial perspective.

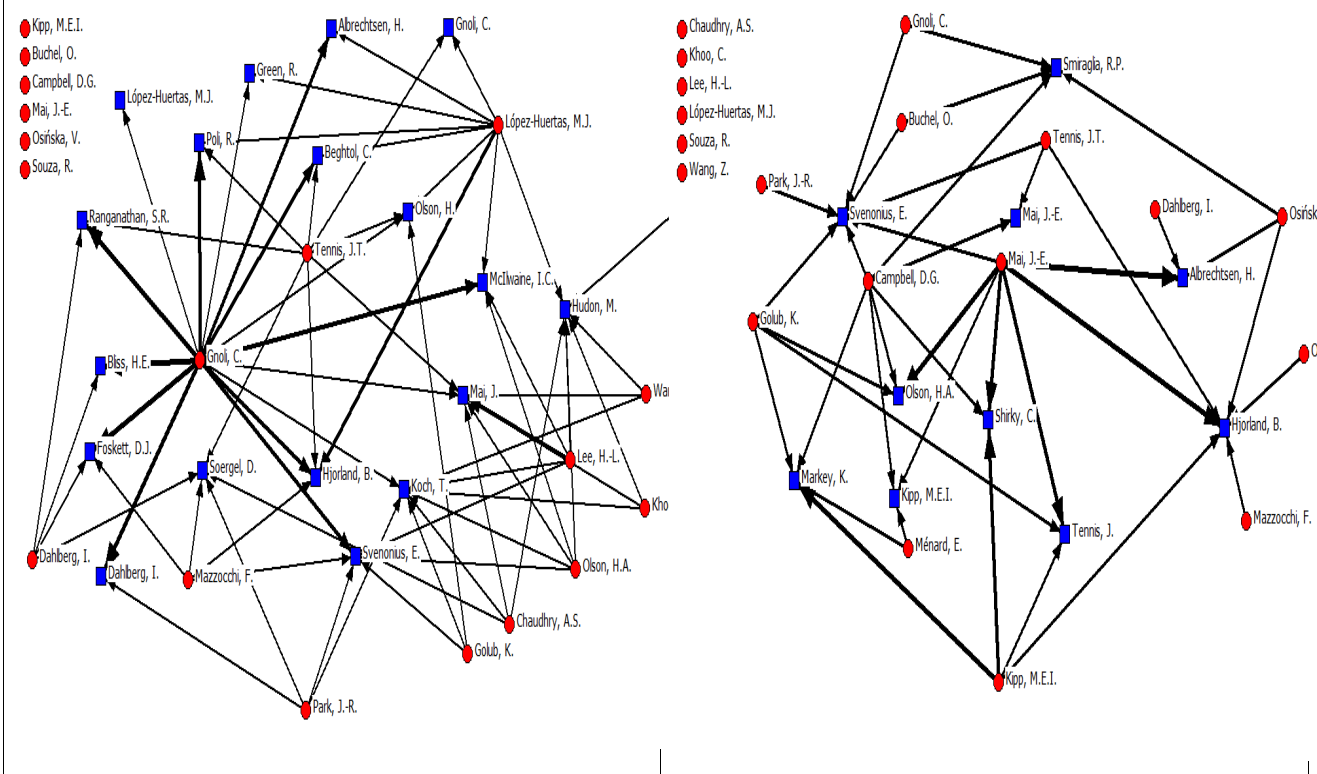
Figure 1a shows the 18 most cited researchers, with at least 4 citations. Researchers Gnoli, Olson and Dahlberg, with, respectively, three, one and one publications, belong also to the group of the most cited ones, with five, four and four citations in the period. In addition, a large number of citing-cited links are observed in this figure, which suggests the plurality and dispersion of theoretical referents adopted in the first five years.

Figure 1b shows the 10 most cited researchers, indicating that in this network, the number of most cited authors is lower than that in Figure 1a, however with a higher frequency of citations, observed by the stronger segments. This behavior indicates that in the second five-year period there is a group of researchers with the highest number of researchers in the area, evidencing a strengthening of theoretical issues, with more specific foundations.

It is observed that Kipp, Mai, Olson and Tennis belong to the group of the most productive and most cited authors, with four, six, five and six citations, respectively.

Figure 1a. Network citing-cited (2004-2008) Figure 1b. Network citing-quoted (2009-2013)

Figure 1a. Citing-cited Network (2004-2008) | Figure 1b. Citing-cited Network (2009-2013)



Source: By the authors

It is also worth noting that, in each five years, six of the most productive authors did not use the most cited authors in their research: in the first five-year period, Kipp, Buchel, Campbell, Mai, Osińska and Souza; and in the second five-year period, Chaudhry, Khoo, Lee, López-Huertas, Wang and Souza are researchers that use different theoretical referents from the remaining 13 researchers in each of the periods. Thus, Souza, the only Brazilian researcher, in neither of the two periods developed his research from the most cited authors.

Also, we highlight the researchers cited in the largest number of articles: in the 2004-2008 period, Koch, cited in 8 articles, Mai in 6 articles, Svenonius and Soergel in 5 articles; and in the 2009-2013 period, Svenonius cited in 7 articles, Hjørland in 6 articles and Smiraglia and Markey in 4 articles.

It should be noted that of the 19 most productive researchers, six are also the most cited ones (~37%), namely: Gnoi, Olson, Dahlberg, Kipp, Mai and Tennis, constituting a group that composes the elite and front of research.

Conclusions

The research showed the researchers, their thematic and theoretical-methodological references used by the journal Knowledge Organization, during the period studied, analyzing the domain in O.C., through the bibliometric approach.

The main themes researched privileged more nuclear subject categories for the development and consolidation of KO.

It should be noted that Svenonius, Albrechtsen, Hjørland, Mai and Olson were cited in the two five-year periods, signaling that classical researchers of the foundations of KO belong to the research front of the scientific elite of the analyzed journal. In addition, Gnoli, Olson and Dahlberg belong both to the research front and to the elite of this journal as they are among the most productive and most cited authors.

Finally, we highlight the greater plurality and dispersion of theoretical referents adopted in the first five years (2004 to 2008) in relation to the following period, in which there is a group of researchers with the highest number of researchers in the area, indicating the strengthening of theoretical issues, .

References

The references was made following the ABNT rules.

ARBOIT, A.E. O processo de (re)construção da teoria do conceito no domínio de Organização do Conhecimento: uma visão dialógica. *Scire*, v. 18, n. 2, p.129-134, 2012.

BARITÉ, M. Organización del Conocimiento: un nuevo marco teórico-conceptual en Bibliotecología y Documentación. In: CARRARA, Kester. (Org.). *Educação, universidade e pesquisa*. Marília, 2001.

ESTEBAN NAVARRO, M. A.; GARCIA MARCO, F. J. Las “Primeras Jornadas sobre Organización del Conocimiento: Organización del Conocimiento e Información Científica”, *Scire*, v.1, n.1, p.149-157, 1995.

GLÄNZEL, W. *Bibliometrics as a research field: a course on theory and application of bibliometric indicators*. Bélgica, 2003.

GUIMARÃES, J. A. C. DOS SANTOS, A. Y., CÂNDIDO, G. G., PINHO, F. A. A dimensão temática da pesquisa em organização do conhecimento: uma análise de domínio dos congressos nacionais e regionais da ISKO. *Scire*, v. 20, n. 2, p.19-25, 2014.

HJØRLAND, B.; ALBRECHTSEN, H. Toward a new horizon in Information Science: domain-analysis. *Journal of the American Society for Information Science*, v. 6, n. 6, p. 400-425, 1995.

HJØRLAND, B. Domain analysis in information science: eleven approaches traditional as well as innovative. *Journal of Documentation*, v. 58, n. 4, p.422-462, 2002

LEYDESDORFF, L. Theories of citation? *Scientometrics*, v. 43, n. 1, 1998.

SMIRAGLIA, R.P. Il congresso brasileiro em representação e organização do conhecimento: Knowledge Organization in Rio 2013 - An editorial. *Knowledge Organization*, v.41, n.2, p.105-112, 2014.

SMIRAGLIA, R.P. ISKO 11's Diverse Bookshelf: an editorial. *Knowledge Organization*, v. 38, n.3, p. 179-186, 2011.

A study on domain in knowledge organization through author citation and co-citation analysis



Lidiane Silva Lima
São Paulo State University
lidyane_lima17@hotmail.com



**Pollyana Ágata Gomes da Rocha
Custódio**
São Paulo State University
pollyanaagata@hotmail.com



Ely Francina Tannuri de Oliveira
São Paulo State University
etannuri@gmail.com



Leilah Santiago Bufrem
São Paulo State University
santiagobufrem@gmail.com

Introduction

Knowledge Organization (KO) studies have played a prominent role in the field of Information Science from its recognition, as an institutionalized science, and have historically fostered the strengthening of the conceptual frameworks of Information Science and Documentation, as they provide the theoretical-methodological bases to better understand the nuclear process of this discipline: mediation between produced knowledge and the socialization of this knowledge.

The research in the area points to at least two important aspects: on the one hand, the theoretical-philosophical approach to epistemological issues by analyzing a domain along with the implications in the organization of classification systems; and,

on the other hand, the methods used in the elaboration of systems in knowledge management (taxonomy, thesauri and ontologies).

Considering the above, this research aims to analyze, through domain analysis in its approach to bibliometric studies, more especially citation and co-citation analysis, the scientific context of "Knowledge Organization" (KO), from the publications indexed in the journal Knowledge Organization (KO), as well as to determine the broad lines that define its scientific community, in its theoretical and methodological references.

This study is justified by the need to broaden and deepen the research in KO, to better signal the current stage of science and the development of the area of "Knowledge Organization" in the international context such as the journal Knowledge Organization.

Author citation and co-citation analysis (ACA)

Citation can be understood as the set of one or more bibliographical references of a publication, which evidences the relations of one publication with another. Citation analysis allows the most influential authors of a scientific domain to be known through the frequency of recorded citations.

Co-citation analysis derived from citation analysis is about the frequency with which two documents, authors or periodicals are cited together in a paper and contributes to the understanding of the intellectual structure of the sciences, evidencing the cognitive and social relations and similarities among these authors, established by the citing community of the studied domain.

According to White and McCain (1998), studies on Author Co-citation Analysis (ACA) had White and Griffith as precursors in 1981, and ACA's main function is to identify the influential authors and show their interrelations, from the recorded citations. In general, researchers with similar research problems cite similar informational sources.

From the construction of ACA networks, it is possible to: (1) identify the group of most co-cited authors; (2) locate this group of authors in their respective schools of research and line of thought; (3) define the degree of centrality of the authors or the central authors and authors conceived as peripheral within the group of this same authors; (4) verify the proximity between the authors and their interrelations and the group of authors that appear interconnected, interfacing several areas of research; and

(5) visualize the authors' structural position in the co-citation network (WHITE; GRIFFITH, 1981, p. 163).

In this context, this research seeks to contribute to the studies of Author Co-citation Analysis (ACA) by analyzing and highlighting the researchers with the highest visibility and their interrelations in Knowledge Organization. It is also considered that mapping scientific networks should be developed in association with epistemological, critical and historical analyzes in order to understand a domain in a fuller and deeper way and to obtain a qualitative analysis.

Methodological procedures

A search on Scopus database was conducted for the title of the journal "Knowledge Organization", filtered in the "source Title" field. Then all the articles published in the journal were retrieved in the period from 2004 to 2013, in the modality article. The selection of the research corpus was conducted out in November 2014 and resulted in a total of 115 articles. The search engine used the keywords: "representação do conhecimento" (knowledge representation), "Organização da informação/conhecimento" (knowledge information/organization), "Ontologias"(ontologies) and "Teorias Classificatórias e catalográficas" (classification and cataloging theories).

A total of 2,884 researchers were identified (without repetitions) for a total of 4,023 references. It was considered those presented five or more citations, in greater number of papers, totalizing 35 cited researchers, based on the assumption that approximately one citation per year would already represent evidence of relevance in the area.

With this data, the occurrence matrix (115x35) was constructed, which refers to the citing and cited authors, in Microsoft Excel software. The square and symmetric matrix of 35x35 size of co-cited authors was generated. The mathematical function "sumproduct function" was used, where the result was the frequency of co-citation among the most cited authors in the 115 analyzed papers.

After this process, Ucinet software was used for the construction of the author co-citation network. For a better visualization of the network, the MDS (Multidimensional Scaling) tool was used as a layout option, which approximates the most similar authors in relation to the co-citation frequencies in a visual arrangement.

The purpose of the layout is to place individuals in the network so that the distance between them reflects their proximity to the co-citation values (OLIVEIRA, 2013).

The obtained data were analyzed and contextualized, following the purpose of delineating the intellectual structure of the area.

Presentation and analysis of results

Table 1 presents the list of 35 researchers cited in the largest number of papers, with at least five citations. In order to account the most influential authors, the self-citations were disregarded, on account of "they do not represent the impact of one work (or author) on another" (FREITAS, 1997, p. 126).

Table 1 - Most cited researchers in the period from 2004 to 2013

RESEARCHERS	# of cited papers
Hjørland, B.	29
Beghtol, C.	20
Olson, H.A.	17
Svenonius, E.	16
Broughton, V.	15
Mai, J.-E.	15
Ranganathan, S.R.	15
Albrechtsen, H.	14
Dahlberg, I.	13
Gnoli, C.	12
Koch, T.	12
Chan, L.M.	11
Zeng, M.L.	11
Vickery, B.C.	11
Bates, M.	10
Jacob, E.K.	10
Chen, H.	9
Vizine-Goetz, D.	9
Foskett, D.J.	8
Lopez-Huertas, M.J.	8
Poli, R.	8
Markey, K.	8
Golub, K.	7
Kipp, M.E.I.	7
Campbell, D.G.	7
McIlwaine, I.C.	7
Smiraglia, R.	7
Buckland, M.	6
Slavic, A.	6
Tennis, J.T.	6
Tudhope, R.	6

Green,R.	5
Hill,L .L.	5
Smith, B.	5
Spink, A.	5

In relation to the most cited researchers, we highlight Hjørland, who presents a thematic scope in his research that reaches the totality of the theoretical lines in Information Science, especially related to Domain Analysis, a methodological approach used in research in the area.

We also note Beghtol, Olson and Svenonius, with 20, 17, 16 cited studies respectively. Beghtol directs his studies to issues related to the global systems of knowledge organization and representation, especially on access, culture and ethics. Olson emphasizes the critical importance of classification in information organization, the use of a medieval language or term on concepts of gender, race, and ethnicity, which can be harmful in the sense that it marginalizes or excludes. Svenonius is widely known for her research on theories of bibliographic control, especially cataloging, classification, and indexing.

Dahlberg, with 13 citations, is a great theoretician of knowledge organization, having developed the Thoery of Concept in the 1970s. Next, Gnoli, the most productive researcher in the period, appears with a total of 12 citations, followed by Koch who focuses his research on the aspects of information theory and digital communication.

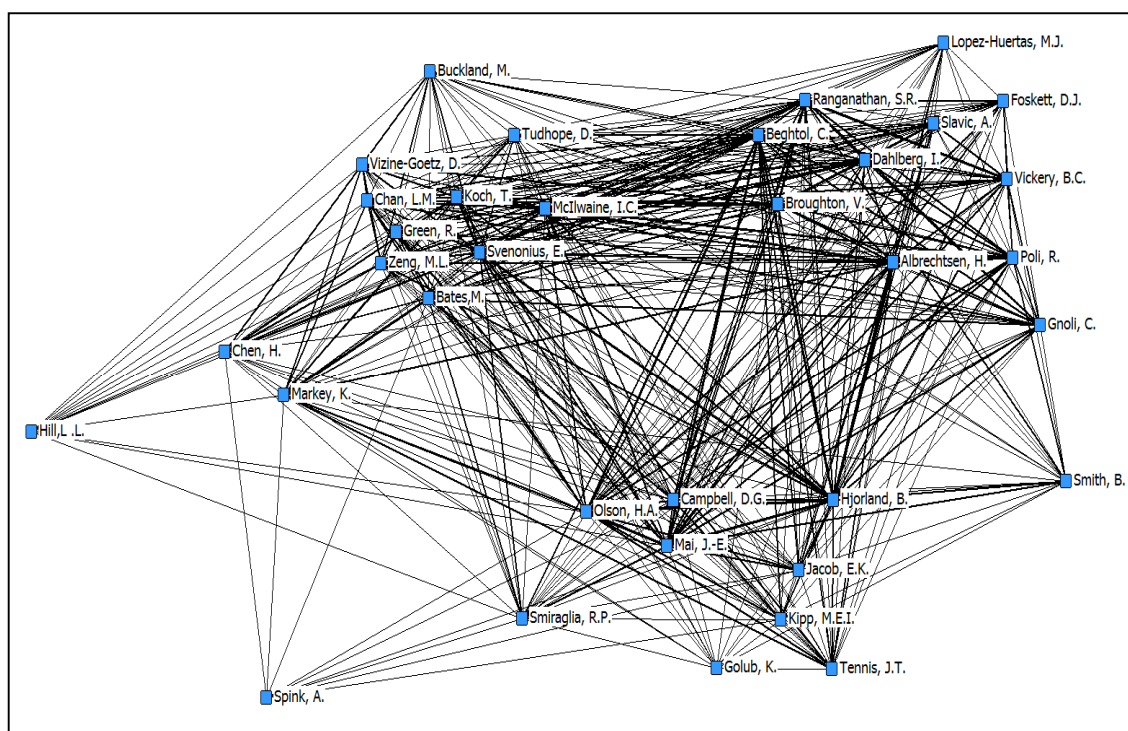
Researchers Chan and Zeng appear with 11 citations each and have developed work on metadata schemes, which can be created or implemented in digital libraries. Bates, cited in 10 papers, contributed to studies in the area of user-centered information system design. Many of his works have been widely cited and used, including articles on his concepts and search tactics for information in the user interface and system.

From the group of the most cited researchers, most are from the area of Information Science. However, researchers with initial training in other areas, mainly Computer Science, used other references to formalize and ground theories about the knowledge organization and information.

The co-citation network, shown in Figure 1, was generated in Ucinet software, where the thickness of the segments represents the co-citation frequency between the

researchers.

Figure 1 - Co-citation network of researchers in Knowledge Organization



Source: by the authors

In relation to the general structure of the network, three large groups stand out, similar in relation to the co-citation frequencies and agglutinated by the MDS. It is observed that the main most cited authors are those found in the axis Hjørland, Albrechtsen, Mai, Beghtol and Dahlberg, whose works form the basis of the main theories in KO, highlighting the importance of the themes Faceted Classification, Theory of Concept, Ontologies and others.

We highlight the authors centered among Hjørland, Mai, Albrechtsen and Beghtol, who together have developed research on Knowledge Organization (KO) and Knowledge Representation (KR).

We observed the lack of Brazilian researchers, however, the researchers who are among the most cited ones are the basis for the theoretical and methodological foundations of the KO research in Brazil.

In spite of setting up a network of a recent period of study, the presence of classical researchers is observed, who not only contributed to the construction of the thematic in Knowledge Organization, but also to the area of Information Science. As an example, the librarian and mathematician Shiyali Ramamrita Ranganathan, more

known for the elaboration of the Five Laws of Library Science (1931), which is considered fundamental to the basis of modern librarianship. In the same line, the Belgian researcher Paul Otlet was considered the father of documentation for establishing classification systems, such as the Universal Decimal Classification (UDC).

Some researchers from technological areas were less prominent, such as Spink, Markey and Hill, who presented systems approaches and online information retrieval in their studies.

Vizine-Goetz, in current research activities, focuses on the application of Functional Requirements for Bibliographic Records (FRBR), models and tools for automated cataloging. She is also a member of the International Federation of Library Associations (IFLA) and ISKO.

Researchers Olson, Beghtol, Campbell, Lopez Huertas and Svenonius are scholars of cultural issues in KO, while Vickery, Foskett, Slavic, Broughton and McIlwaine represent the group of English scholars linked to classification theories. The dyad of researchers Gnoli and Poli developed studies on ontologies. They attribute to Dahlberg the role of structuring source in bibliographic classification.

Svenonius, Zeng, Bates, and Green represent an American group linked to the theory of subject indexing and cataloging and to the Library of Congress Classification (LCC) and CDD classification systems.

It is possible to observe a group of scholars involved with the epistemological issues of KO, philosophy of science and also of studies about the new paradigms in the context of Information Science represented by Hjørland, Jacob, Mai, Tennis, Smiraglia and Gnoli. In this group, Hjørland, Mai, Smiraglia and Tennis have studies focused on domain analysis methodologies.

Conclusions

This research focused on identifying the most cited and co-cited authors in the scientific production in "Knowledge Organization", in the journal Knowledge Organization, from 2004 to 2013, in order to characterize its domain.

From the most cited group of authors, most are researchers with consistent research in Information Science. The co-citations highlighted by the strongest segments in the network show the researchers, the themes, the discursive community and the social and cognitive structure of the scientific domain.

However, there is a lack of well-known and prominent KO researchers among the most cited ones, such as Lancaster and Bliss, who, at different periods, contributed to studies and research that provided the basis for the construction of theory and practice in indexing and summaries: theory and practice, the greatest contribution of the author occurs within the scope of classification theory and systemic theory.

References

The references was made following the ABNT rules.

FREITAS, M. H. de A. Oito anos de Transinformação. *Transinformação*, Campinas, v. 9, n. 3, set./dez., 1997.

OLIVEIRA, E. F. T. de. *Análise de Domínio em “Estudos Métricos” no Brasil: produção, impacto e visibilidade em âmbito nacional e internacional*. 2013. 193 f. Tese (Livre-Docência em Estudos Métricos em Produção e Organização da Informação). Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2013.

WHITE, H, H.D.; GRIFFITH, B. Author cocitation; a literature measure of intellectual structure. *Journal of the American Society for Information Science & Technology*, v. 32, n. 2, p. 163-171, 1981.

WHITE, H.D.; MCCAIN, K.W. Visualizing a discipline: an author co-citation analysis of Information Science, 1972-1995. *Journal of the American Society for Information Science*, v. 49, n. 4, p. 327-355, 1998.

Classification approaches of the papers submitted to ENANCIB WG2: professional classification and non-professional classification



Walter Moreira
São Paulo State University
walter.moreira@marilia.unesp.br



Isabela Santana de Moraes
São Paulo State University
isabela@marilia.unesp.br

Introduction

The Classification Theory, as a guiding and questioning principle of the knowledge organization and representation processes, has been addressed in several papers presented in scientific conferences in information science. Although the thematic of these events may vary, the problematic regarding the classification, even with other names, is a striking presence. This is due to the protagonism of the action of classification in the cognitive process, in general, and in knowledge organization, especially in libraries.

Thus, observing the polysemy of the expression "classification", this study aims, generally, to understand the concepts of "professional classification" and "non-professional classification". In this case, terminology is also the subject of investigation. Regardless, however, of the terminological consensus, there are two different meanings of classification that are effectively and dialectically complementary. We adopt the point of view of Information Science for the dichotomy "professional classification" and "non-professional classification". Thus, the classification developed by a veterinary researcher, for example, to organize the concepts, objects or beings in his/her field of study, will be taken as a non-professional classification. It is assumed that, in this case, this researcher first aims at the discovery of knowledge, without the necessary intentions of theoretical generalizations regarding classification as a

process or as a product. In these cases, classification theory, when observed, is used in its purely applied character, that is, this classification does not aim at the study of theory for the advance of the theory itself.

The objectives of this paper are: a) to discuss the concepts of professional and non-professional classification in the light of the classification theory; and b) to identify and describe the presence of this theme in the Brazilian literature of information science.

The papers presented to the Working Group " Knowledge Organization and Representation" (WG02) published in the proceedings of the National Meeting of Research in Information Science (ENANCIB), with retrospective five-year period (2010 to 2014) was used as the corpus of the study.

ENANCIB is an annual scientific meeting promoted by Associação Nacional de Pesquisa e Pós-graduação em Ciência da Informação (ANCIB) and is one of the most relevant and significant spaces for the discussion and consolidation of information science in Brazil. The meeting is thematically organized in eleven working groups.

The term "classific" was used as a search strategy, so that the term "classification" could be located in its variations of number and in its adjectival forms, such as "bibliographic classification", "documentary classification", and so on. The search was performed directly in the documentary corpus, that is, in the set of papers presented to WG02 in the established period, using automatic resources to identify occurrences of the search expression directly in the files. The documentary corpus consisted of 187 papers, with the following annual distribution: 2010 (n=29); 2011 (n=29); 2012 (n=34); 2013 (n=40) and 2014 (n=55).

For the composition of the corpus of analysis, we observed the presence, simultaneous or not, of the term "classification" in the fields Title and Keywords. These fields were chosen because of their greater representational power. Following this criterion, 29 papers were found.

After exploratory reading, in order to understand the general theme of the paper, the set of bibliographic references was analyzed, to determine the occurrence of two categories of analysis: a) classic reference of the classification theory/knowledge organization and representation (CT/KOR) and b) referential of "other" fields or subfields.

The following sections aim to relate bibliographic classification to the major field of knowledge organization and representation (section 2), as well as to present its

theoretical underpinnings (section 3) and conceptual bases (section 4). This brief theoretical-conceptual foundation establishes the conceptual bases for the analysis and discussion of the results, object of section 5. The conclusions of the research are presented in section 6.

Classification in knowledge organization and representation

The field of knowledge organization and representation (KOR) is structured in two actions (organize and represent) around an object (knowledge). These actions are operationalized using tools, processes and products (FUJITA, 2008). Also observing its theoretical-conceptual character, KOR takes care of the epistemological bases underlying these actions and is concerned with the development, application and evaluation of knowledge organization systems.

Given the breadth of the object (knowledge), it is necessary to observe the complementary meanings that can be attributed to knowledge organization (HJORLAND, 2003; 2008). In a broad sense, it is related to the structural and social organization of knowledge in order to meet the system of disciplines and professions. In the strict sense, it includes activities such as "[...] document description, indexing and classification performed in libraries, bibliographical databases, archives and other kinds of 'memory institutions', by librarians, archivists, information specialists, subject specialists [...]" (HJORLAND, 2008, p.86).

The theoretical-practical activities of knowledge organization and representation, such as classification and indexing, are thus strongly connected as the social organization of knowledge. There would be no point in maintaining "memory institutions" that did not dialogue with what Dahlberg called "knowledge in action" (DAHLBERG, 1993; GUIMARÃES, 2008).

Classification Theory

The classification process is pointed out as a necessary and indispensable resource to man in his relation of knowledge to the world that surrounds him (BROUGHTON, 2004; BATLEY, 2005) and often manifests itself in an unconscious way. For Jacob (2010), it is not the classification, but the cognitive process of categorization that is ubiquitous. The distinction is still not peaceful and the concepts of categorization and classification actually have many similarity traits (JACOB, 2004).

By adopting the perspective that knowing is to establish relationships among concepts, it is necessary, therefore, as a cognitive task, that such concepts be identified in their characteristics and thus observed with regard to their similarities and differences. Thus, the identification of concept occurs through classification processes that evidence the contrast or the identity among concepts. Classification, in this sense, refers to the grouping of things that are similar and concerns the imposition of some kind of structure on the understanding of reality (BATLEY, 2005).

This structure (referring to common sense) is also rebutted in a representation model, even if not formalized and yet elaborated ad hoc, only for the understanding of a certain fact or phenomenon, in a level of reduced conceptual sharing. In the case of bibliographic classification systems, in which the representational character is more evident and necessary to its pragmatic aspect, orientated to the construction of knowledge representations and of informational objects, the representation requires more elaborated levels of consensual formalization that meet the interests of the system as a whole, including classifiers and users.

In classification systems, the notion of grouping refers to the idea of division or separation by similarities and differences; ordering, in turn, is related to the concept of assortment, in the sense of mixing and combining different things. The notion of ordering also refers to the idea of symbolic representation of elements and the sequence they hold. The representation, in this case, fulfills two functions: it allows the element removed from its position to be replaced and it allows the identification of the "correct" place of insertion of a new element, preserving the logic of the system (BROUGHTON, 2004).

The history of the classification theory can be punctuated by the pioneerism of some projects whose elaboration processes act as theoretical-conceptual frameworks, among them Dewey's Decimal Classification, published in 1876, and Otlet and La Fontaine's Universal Decimal Classification, in 1905. These systems have been updated and evolved over time (DDC, for example, is in its 23rd edition), and are still widely used, despite some limitations.

In the 1930s, Bliss's (Bibliographic Classification 1 and later Bibliographic Classification 2) and Ranganathan's (Colon Classification) contributions modernized classification theory with the facet approach in analytic-synthetic systems. This occurred by more flexible and more accurate identification of relationships among the

classes, the use of more synthetic notations, the use of facets, and the consistent application of the citation order.

The faceted approach to classification was broadened in the 1950s by the work of the Classification Research Group (CRG) and influenced a number of other classification systems (including DDC and UDC) by flexibilizing them and giving them an analytical-synthetic character, in opposition to the more rigid characteristic of purely enumerative systems.

The concept of classification

At first, it is necessary to understand that all classification is arbitrary and must be observed in its relativization character. The expression "arbitrary" is ambiguous in Brazilian-Portuguese language and allows at least three uses, according to Aulete Dictionary: "1. which depends on the agency or will of the one who decides; which has no established rules (arbitrary measures); 2. which follows one's will, regardless of the opinion or need of others; DESPOTIC; 3. not necessary or required; OPTIONAL" (LACERDA, 2015). The meaning to be given to arbitrariness in this context is supported by Foucault's (2000) understanding of this term, which is closer to meanings 1 and 2 in the dictionary:

[...] one cannot know the order of things 'in their isolated nature, but by discovering that which is the simplest, then that which is the next simplest, one can progress inevitably to the most complex things of all. (FOUCAULT, 2000, p. 73).

[...] a thing can be absolute according to one relation yet relative according to others; order can be at once necessary and natural (in relation to thought) and arbitrary (in relation to things), since, according to the way we consider it, the same thing may be placed at differing points in our order (FOUCAULT, 2000, p. 74)

The arbitrary character of classification, as seen, does not exempt it from presenting, at any level and for any purpose, logical interconsistency among all its components.

In the search for theoretical and methodological understanding of classification, Beghtol published an article in the journal *Knowledge Organization* in 2003 comparing some aspects of "professional classification" and "naïve classification". The text gave rise to a heated debate in the same journal: Hjørland and Nicolaisen (2003) reacted to the use of the expression naïve. Beghtol (2004) clarified some aspects and the text was also rejoiced by Nicolaisen and Hjørland (2004). Six years later the debate was still shaking with the work of Jacob (2010).

Beghtol (2003; 2004) names the "professional" classification oriented to information retrieval and the "naïve" classification as the one made for the purpose of knowledge discovery. According to this distinction, classification for the purpose of information retrieval (professional) is based on consensus and literary warrant, based on more stable and widely shared conceptions. Exactly why their reviews are more time consuming. The classification for knowledge discovery (naïve), in turn, is related to the discovery of the new, and for this it uses the new as a resource - and more provisional than the previous one - classification scheme, that is, a new organization hypothesis. The methods used by both professional classification and naïve classification are similar, according to Beghtol (2003).

The scientific and academic classifications, Hjørland and Nicolaisen (2004) reacted, are not naïve. The authors consider that Beghtol's understanding of the relations between academic and scientific classification and the classification made by librarianship and information science were misleading. The bibliographic classification, they argue, is largely dependent on academic and scientific classifications, even if they are not given due attention, as they say: "We in LIS should obviously be concerned with generalized principles and methods of classification. Very often it seems, however, as if we ignore the work done by scientists, philosophers and scholars" (HJORLAND; NICOLAISEN, 2004, p 57).

Beghtol's theme (2003, 2004), some texts that gave rise to it (DAVIES, 1989, KWASNIK, 1999; RAFFERTY, 2001), as well as the issues raised in the texts that debated his proposal (HJORLAND; NICOLAISEN, 2004; NICOLAISEN; HJORLAND, 2004; JACOB, 2010, KWASNIK; FLAHERTY, 2010) motivated the proposal of this investigation in order to explore this work (with further study) and verify how KOR in Brazil has understood the relationship between "professional classification" and the "non-professional classification", observing, in addition, the need for investment in the conceptual analysis of these terms in the area of information science.

Results and discussion

The distinction "CT/KOR reference" and "others reference", as proposed in the objectives of this work, should not be taken as evidence of demarcation of field exclusivity, domain reservation or any demonstration of disciplinary xenophobia. This dichotomy serves only to mark the analysis and to verify the directions that the research on classification follows in the corpus of analysis. In this way, one must consider the subjectivity and the limits of this type of categorization.

The category "others" refers to texts that are not classically identified with CT/KOR, but are not necessarily foreign. It includes, for example, seminal texts of information science and, in this condition, dialogue with the CT/KOR reference. In a way, in fact, all the texts cited by the authors in their papers are aimed at promoting dialogue with the CT/KOR.

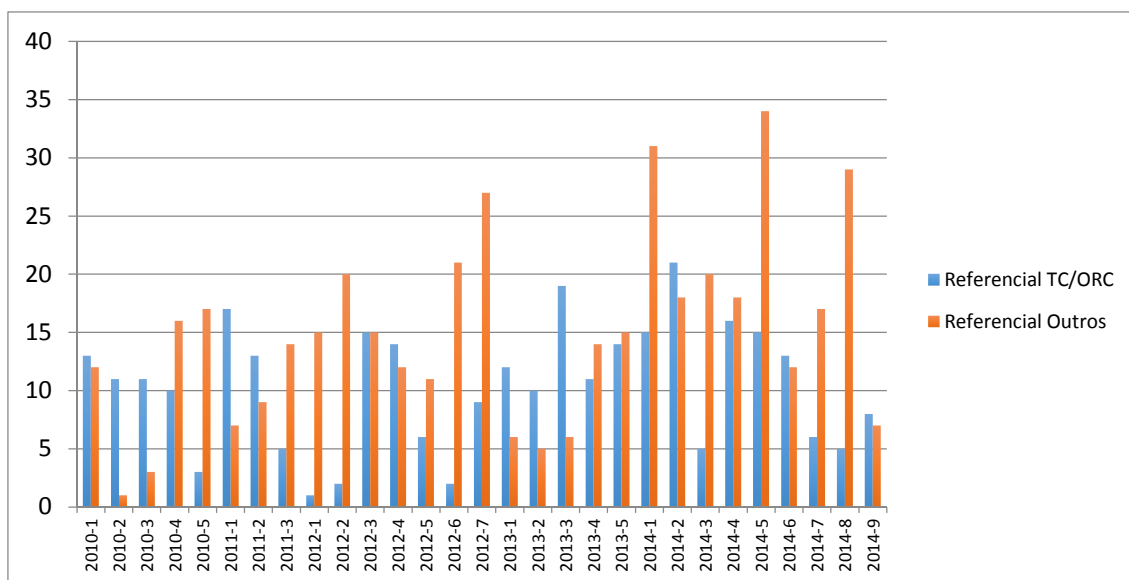
For the process of references categorization, we observed: the title, the nature of the item (book, journal article, papers published in conferences, etc.) and the authors' affiliation. As examples, by the number of occurrences, we can mention: *Teoria da classificação*, by Requião Piedade, as representative of CT/KOR and *Métodos e técnicas de pesquisa social*, by Antonio Carlos Gil, as representative of "others".

The total number of analyzed articles was 29, with the following annual distribution: 2010=5; 2011=3; 2012=7; 2013=5; 2014=9. Figure 1 shows the general synthesis of the data.

The average number of references per article is 25.3, with 25 as the fashion number. This set of references is distributed as follows, on average: 10.4 CT/KOR references per work and 14.9 "others" references.

In the general set of data, there is a predominance of "others" reference which may reveal some tendency to be verified in further studies. Considering the total number of references used in the 29 studies ($n = 734$), the CT/KOR subtotal represents 41.1% of the total ($n = 302$) and the subtotal of references "other" equals to 58.9 % from the total ($n = 432$).

Figure 8 - Use of the theoretical CT/KOR reference and "others" in the corpus of analysis



The five papers with a higher incidence of "other" references address topics of classification that are closer to the non-professional classification, according to the concept presented previously, and refer to the analysis and development of classification proposals in the field of music, communication and science and technology (Table 1).

Table 1 - Papers with the highest percentage of "others" reference

Paper	Focus	CT/KOR References	"Others" References	% others	Total references
2012-1	music	1	16	93,8	17
2012-6	science and technology	2	21	91,3	23
2012-2	communication	2	20	90,9	22
2014-8	science, technology and innovation	5	29	85,3	34
2010-5	communication	3	17	85	20

Source: Research data, 2015

Such studies, when brought to the discussion at a meeting in the area of information science, are no longer fully identified with the concept of professional classification in the strict sense. We observed the origin of the problems that motivated these studies for the purpose of ordering these works, in addition to the used references.

Conclusions

The approach to the problem, observing the limits of this work, is exploratory and still needs to be expanded so that more precise interpretations can be produced. There are, however, some quantitative aspects that seem to confirm a greater approximation between the theme of non-professional classification and the use of a complementary reference to a more circumscribed reference to the classification theory.

It is also observed (Figure 1) that there is a growing tendency for the dialogue between classification theory and the bibliographic reference of other areas or other subfields of information science. A more accurate observation of this phenomenon can help in the terminological clarification of the concepts of classification, professional or not, and can help to clarify the differences in purposes and the relationships between the different classification modalities, surpassing, more effectively, the limits of the classification theory within knowledge organization and representation.

In general, it is believed that this type of approach contributes to avoiding reductionist approaches to bibliographic classification.

References

The references was made following the ABNT rules.

BATLEY, Sue. *Classification in theory and practice*. Oxford: Chandos, 2005.

BEGHTOL, Clare. Classification for information retrieval and classification for knowledge discovery: relationships between "professional" and "naïve" classifications. *Knowledge organization*, v. 30, n. 2, p. 64-73, 2003.

_____. Response to Hjørland and Nicolaisen. *Knowledge Organization*, v. 31, n. 1, p. 62-63, 2004.

BROUGHTON, Vanda. *Essential classification*. New York: Neal-Schuman, 2004.

DAHLBERG, Ingetraut. Knowledge organization: its scopes and possibilities. *Knowledge organization*, n. 20, p. 211-222, 1993.

DAVIES, Roy. The creation of new knowledge by information retrieval and classification. *Journal of Documentation*, v. 45, n. 4, p. 273-301, 1989.

FOUCAULT, Michel. *As palavras e as coisas: uma arqueologia das ciências humanas*. São Paulo: Martins Fontes, 2000.

FUJITA, M. S. L. Organização e representação do conhecimento no Brasil: análise de aspectos conceituais e da produção científica do ENANCIB no período de 2005 a 2007. *Tendências da Pesquisa Brasileira em Ciência da Informação*, v. 1, p. 1-32, 2008.

GUIMARÃES, José Augusto C. A dimensão teórica do tratamento temático da informação e suas interlocuções com o universo científico da International Society for Knowledge Organization (ISKO). *Revista Ibero-americana de Ciência da Informação (RICI)*, v.1 n.1, p.77-99, jan./jun. 2008.

HJORLAND, Birger; NICOLAISEN, Jeppe. Scientific and scholarly classifications are not "naïve": a comment to Beghtol(2003). *Knowledge Organization*, v. 31, n. 1, p. 55-63, 2004.

HJORLAND, Birger. Fundamentals of knowledge organization. *Knowledge Organization*, v. 30, n. 2, p. 87-111, 2003.

_____. What is knowledge organization (KO)? *Knowledge Organization*, v. 35, n. 2/3, p.86-101, 2008.

JACOB, Elin K. Classification and categorization: a difference that makes a difference. *Library Trends*, v. 52, n. 3, p. 515-540, winter 2004.

_____. Proposal for a classification of classifications built on Beghtol's distinction between "naïve classification" and "professional classification". *Knowledge Organization*, v. 37, n. 2, p. 111-120, 2010.

KWASNIK, Barbara H. The role of classification in knowledge representation and discovery. *Library Trends*, v. 48, n. 1, p. 22-47, 1999.

_____.; FLAHERTY, Mary Grace. Harmonizing professional and non-professional classifications for enhanced knowledge representation. In: GNOLI, Claudio; MAZZOCCHI, F. (Eds.). *Paradigms and conceptual systems in knowledge organization: proceedings of the 11th International ISKO Conference, 23-26 feb. 2010, Rome, Italy*. Würzburg: Ergon-Verlag, 2010. p. 229-235.

LACERDA, Carlos Augusto (Ed.) *iDicionário Aulete*. [s.l.]: Lexicon Editora Digital, 2015. Disponível em: <<http://www.aulete.com.br/arbitrário>>. Acesso em: 27 maio 2015.

NICOLAISEN, Jeppe; HJORLAND, Birger. A rejoinder to Beghtol. *Knowledge organization*, v. 31, n. 3, p. 199-201, 2004.

RAFFERTY, P. The representation of knowledge in library classification schemes. *Knowledge organization*, v.28, n.4, p.180-191, 2001

Information behavior within ISKO - International Society For Knowledge Organization: a study for the 2004-2014 period



Marli Vitor da Silva
São Paulo State University
marli_biblio@hotmail.com



Helen de Castro Silva Casarin
São Paulo State University
helen-cs@uol.com.br

Introduction

The study on the information behavior of individuals and the use they make of the obtained information is essential for the foundation of Information Science as several issues in the area, such as information organization and representation of documents are subordinate to it. (ALBRECHTSEN & HJØRLAND, 1997).

Hjørland (1994) points out that people as information users are also information producers. Information production is also relevant to studies related to the area of information organization. In this way, it is possible to consider studies on information behavior relevant to the area of information organization as a sub-area of Information Science. From this assumption, one can ask the following question: how do the area studies, in particular those presented at ISKO meetings, have addressed the issue of informational behavior?

Thus, this paper aimed to verify how the theme information behavior has been discussed and treated in the literature on information organization, specifically within the scope of the International Society for Knowledge Organization. This meeting was chosen as it is one of the most important in the area of information organization and addresses the main issues discussed by the academic community regarding this topic.

Information organization and Information behavior

Information organization is a process that concerns the physical and content

description of informational resources, both traditional and electronic (BRÄSCHER; CAFÉ, 2008).

It should be noted that, according to Garcia Marco (1995), a comprehensive theory of information organization must be constructed from references of different areas, including "Studying the characteristics of specialized discourses in general, and scientists in particular" (BARITÉ, 2001, p.57). Thus, it is noted that the study of the characteristics and behavior of groups of individuals from different areas of knowledge that make up discursive communities are relevant to information organization.

It is important that information organization methods take into account users' behavior for information seeking and use in different contexts and knowledge domains (HJØRLAND, 1994 e 2005), since the "[...] tools, concepts, meaning, information structures, languages for special purposes (LSP), information needs, and relevance criteria are shaped in discourse communities, for example, in scientific disciplines, which are parts of society's division of labor" (HJØRLAND, B., ANDERSEN, J., & SØNDERGAARD, 2005, p.258). This approach is important as a methodological contribution as it allows to characterize and evaluate a particular science, since it allows the identification of the conditions by which scientific knowledge is built and socialized (GUIMARÃES, 2015).

According to Hjørland (1997), users' behavior should be analyzed from the perspective of a scientific situation in a given area, since the investigation of "micro events" of information seeking behavior does not bring significant results to the area of information organization as they do not reflect the reality about the seeking behavior of users belonging to a certain group (to a discursive community). For the author, it is not possible to provide users with access to information sources without any knowledge of what is relevant or not for a given community.

Hjørland (2002, p. 432) considers that the studies on information users are fundamental to the analysis of a given domain and its characterization, as they allow to identify the informational needs of individuals in different communities, therefore, "[...] that a domain is organized according to its users' preferences, behaviors or mental models" (GUIMARÃES, 2015, p.5).

Wilson (2000, p. 49) defines Information Behavior as "[...] the totality of human behavior in relation to sources and channels of information, including both active and passive information seeking, and information use".

According to Wilson (2000), the term information behavior encompasses two others, which complement one another: a) Information Seeking Behavior, which refers to the different procedures adopted by individuals to locate and access information to meet their Information needs, and (b) Information Search Behavior, which is related to users' interactions with computer-based information systems.

Methodology

In order to meet the objective of the proposed study, domain analysis was used. In order to identify the studies of information organization that address the subject of information behavior, the proceedings ISKO's (International Society for Knowledge Organization international) conferences were consulted.

The 2004-2014 period was chosen for the study, which includes the conferences held in London (England), Vienna (Austria), Montreal (Canada), Rome (Italy), Mysore (India) and Krakow (Poland). Data collection was carried out by consulting the printed records (2004 and 2014) and by searching the digital records (2006, 2008, 2010 and 2012). The searches were performed using the keywords: information behavior (our); information seeking behavior (our); information search behavior (our). First, the title and the abstract of the papers were verified. When the paper appeared to be relevant to the study, the content of the text was checked in its entirety.

The analysis of the selected papers was performed through Content Analysis (BARDIN, 2010). Thus, the following categories of analysis were established: incidence of papers on informational behavior; methodology; scope of the study (information behavior, information seeking behavior and search behavior in information systems).

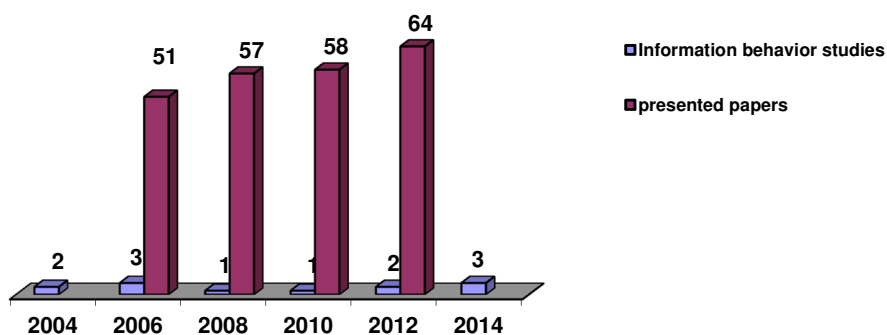
Presentation and discussion of results

The total of papers collected in the proceedings of ISKO International editions in the period covered in this study (2004-2014) was 230 papers. From this total, 12 papers were analyzed as they address information behavior, that is, 5.2% of the total, revealing that this theme is little approached in the studies of the area.

The incidence of studies over the studied period, although small, has been constant throughout the studied period, varying between two and three papers at each

edition of the conference, with the exception of the editions of 2008 and 2010, in which only one paper was verified. These data can be visualized in Graph 1.

Graph 1: Incidence of papers on information behavior in the proceedings of ISKO International (2004-2014)



Thus, it can be affirmed that the incidence of papers presented on the subject of informational behavior, in comparison to the total of papers presented at the conferences does not seem very significant, although it is constant.

Regarding the origin of the authors of the selected papers, it was verified that they come from several countries. Among these, the United States was the most prominent, as six out of the 12 identified papers are by U.S. authors (MILLER (2004), PAJARILLO, (2006); LEE, (2006); KIPP, (2008); TILLEY, LA BARRE (2010); TILLEY, LA BARRE, (2012). The other countries were between two presentations (MATERSKA (2014), KAMINSKA, PULAK (2014) from Poland) or only one presentation (SHIRI, REVIE (2004) from England), (WILLIAMSON (2006) from Canada), (BERNAOUI, HASSOUN (2012) from Nigeria), AKERELE, DAVID, OSOFIAN (2014) from France).

It is noteworthy that two American researchers, Kathryn A. La Barre and Carol L. Tilley, were the most outstanding, presenting papers on the theme in two editions of the event, namely: Rome (2010) and Mysore (2012).

<i>Type of study</i>	<i>Articles</i>
<i>Information behavior (our)</i>	PAJARILLO, Edmund JY. A qualitative research on the use of knowledge organization in nursing information behavior.

Information seeking behavior (our)	TILLEY, Carol L.; LA BARRE, Kathryn A. New models from old tools Leveraging an understanding of information tasks and subject domain to support enhanced discovery and access to folktales
	TILLEY, Carol L. LA BARRE, Kathryn A. What if they build it and no one comes? Balancing Full-Text Access and User Tasks
	BERNAOUI, Radia; HASSOUN, Mohamed. User Expectations, Reality and Delineation of Agricultural Information Systems in the Maghreb
	KAMINSKA, Aneta; PULAK, Irena. Knowledge organization in a digital learning environment in the experiences of pedagogy students
Information search behavior (our)	SHIRI, Ali Asghar; REVIE, Crawford; End-user interaction with thesauri: an evaluation of cognitive overlap in search term selection.
	MILLER, Danielle H. User perception and online online catalogue: public library OPAC users "think aloud".
	WILLIAMSON, Nancy J. Knowledge Structures and the Internet: Progress and Prospects
	LEE, Hur-Li Navigating Hierarchies vs. Searching by Keyword: Two Cultural Perspectives
	KIPP, Margaret E. I. Searching with Tags: Do Tags Help Users Find Things?
	AKERELE, Olubunmi; DAVID, Amos; OSOFISAN, Adenike. Using the concepts of case based reasoning and basic categories for enhancing adaptation to the user's level of knowledge in decision support system
	MATERSKA, Katarzyna. Faceted navigation in search and discovery tools

Source: The authors

It is noteworthy that four out of the 12 identified papers used the combination of different methods of data collection in their research, that is: PAJARILLO, Edmund J. Y. (2006) - Semi-structured interview and Focus Group; MILLER, Danielle H. (2004) - Questionnaire and Search Task; LEE, Hur-Li (2006) - Semi-structured interview and Search Task e TILLEY, Carol L.; LA BARRE, Kathryn A. (2010) - Semi-structured interview and Facet Analysis.

As showed at Chart 1, seven articles can be classified as *Information seeking behavior* and four as *Information search behavior*. Only one can be classified as *Information behavior* study, that is the broader field of investigation that include the others. According to Wilson's (1999) definition about the type of studies concerning users behavior, the main articles studies can be categorized in the micro level of Information behavior, that is, information seeking and information search.

The methodologies used for data collection in the papers did not vary much. The most used methods were: experimental, in which subjects were asked to perform information search tasks (SHIRI; REVIE, 2004; MILLER, 2004; WILLIAMSON, 2006; LEE, 2006; KIPP, 2008); semi-structured interview (PAJARILLO, 2006; TILLEY, LABARRE, 2010; TILLEY; LABARRE, 2012); and questionnaire (MILLER, 2004; BERNAOUI, HASSOUN, 2012 and KAMINSK; PULAK, 2014).

Conclusions

The present study revealed that only 27.6% of the studies address information behavior, that is, present the focus on the user, which demonstrates a lack of research on this subject in the area.

The collected data also show that, although few, there is a certain regularity in the production of research on the subject information behavior, since in the studied period the theme was always present in at least one paper.

The results also revealed that the most studied issues are those related to the behavior of individuals when they are actively seeking information (Information Seeking Behavior), or when they are in direct contact with information systems (Information Search Behavior), both classified as micro-. These studies are aimed both at improving the systems and at improving the techniques used by individuals in retrieving relevant information. The thematic Information Behavior, which is more comprehensive and includes search behavior and search behavior in information systems, has been little explored by the area of information organization, requiring more research in the area.

References

The references was made following the ABNT rules.

AKERELE, O.; DAVID, A.; OSOFISAN, A. Using the concepts of case based reasoning and basic categories for enhancing adaptation to the user's level of knowledge in decision support system. In: *Thirteenth International ISKO Conference - Knowledge organization in the 21st century: between historical patterns and future prospects*, 2014, Kraków, Poland, May 19-22. Proceedings of the Thirteenth International ISKO Conference. Würzburg: Ergon, 2014.

ALBRECHTSEN, H.; HJÖRLAND, B. Information seeking and knowledge organization. *Knowledge Organization*, v. 24, n. 3, p.136-144, 1997.

BRÄSCHER, M.; CAFÉ, L. *Organização da Informação ou Organização do Conhecimento*, 2008. Disponível em: <<http://www.ancib.org.br/media/dissertacao/1835.pdf>>. Acesso em: 08 abr. 2014.

BARDIN, L. *Análise de conteúdo*. Lisboa: Edições 70, 2010. 281 p.

BARITÉ, M. Organización Del conocimiento: un nuevo marco teórico-conceptual en Bibliotecología y Documentación. In: CARRARA, K. (Org.). *Educação, universidade e pesquisa*. Marília: Unesp – Marília – Publicações, 2001. p. 35-60.

BERNAOUI, R.; HASSOUN, M. User Expectations, Reality and Delineation of Agricultural Information Systems in the Maghreb. In: Twelfth International ISKO Conference - Categories, contexts and relations in knowledge organization, 2012, Mysoure, India, August 6-9, 2012. *Proceedings of the Twelfth International ISKO Conference*. Würzburg: Ergon, 2012.

GARCIA MARCO, J. Los contenidos y la secuencia docente de la Organización y Representación del Conocimiento: una propuesta interdisciplinar. *Organización del Conocimiento en Sit. Inf. Y Docum.*, v. 1, p. 219-228, 1995.

GUIMARÃES, J. A. C. *A análise de domínio com perspectiva metodológica em organização da informação*, 2015, 12 p.

HJØRLAND, B., ANDERSEN, J., & SØNDERGAARD, T. F. UNISIST model and knowledge domains. In: *ENCYCLOPEDIA of library and information science: First Update Supplement*. New York: C R C Press LLC, 2005. p. 129 - 135.

HJORLAND, B. Nine Principles of knowledge organization. *Advances in knowledge organization*, v. 4, p. 91-100, 1994.

HJORLAND, B. *Information Seeking and subject representation: an activity-theoretical approach to Information Science*. London: Greenwood Press, 1997. 214 p.

HJORLAND, B. *Journal of the American Society for Information Science and Technology*, v. 53, n. 4, p. 257-270, 2002.

KAMINSKA, A.; PULAK, I. Knowledge organization in a digital learning environment in the experiences of pedagogy students. In: Thirteenth International ISKO Conference - Knowledge organization in the 21st century: between historical patterns and future prospects, 2014, Kraków, Poland, May 19-22. *Proceedings of the Thirteenth International ISKO Conference*. Würzburg: Ergon, 2014.

KIPP, M. E. I. Searching with Tags: Do Tags Help Users Find Things? In: tenth International ISKO Conference - Culture and identity in knowledge organization, 2008, Montréal, Canada, August 5-8. *Proceedings of the Tenth International ISKO Conference*. Würzburg: Ergon, 2008.

LEE, Hur-Li Navigating Hierarchies vs. Searching by Keyword: Two Cultural Perspectives. In: Ninth International ISKO Conference - Knowledge organization for a global learning societ, 2006, Vienna, Austria, July 4-7. *Proceedings of the Ninth International ISKO Conference*. Würzburg: Ergon, 2006.

MATERSKA, K. Faceted navigation in search and discovery tools. In: Thirteenth International ISKO Conference - Knowledge organization in the 21st century: between historical patterns and future prospects, 2014, Kraków, Poland, May 19-22. *Proceedings of the Thirteenth International ISKO Conference*. Würzburg: Ergon, 2014.

MILLER, D. H. User perception and online online catalogue: public library OPAC users “think aloud”. In: Eighth International ISKO Conference - Knowledge organization and the global information society, 2004, Londres, England, July 13-16. *Proceedings of the Eighth International ISKO Conference*. Würzburg: Ergon, 2004.

PAJARILLO, E. JY. A qualitative research on the use of knowledge organization in nursing information behavior. In: Ninth International ISKO Conference - Knowledge organization for a global learning societ, 2006, Vienna, Austria, July 4-7. *Proceedings of the Ninth International ISKO Conference*. Würzburg: Ergon, 2006.

SHIRI, A. A.; REVIE, C. End-user interaction with thesauri: an evaluation of cognitive overlap in search term selecton. In: Eighth International ISKO Conference - Knowledge organization and the global information society, 2004, Londres, England, July 13-16. *Proceedings of the Eighth International ISKO Conference*. Würzburg: Ergon, 2004.

TILLEY, C. L.; LA BARRE, K. A. New models from old tools Leveraging an understanding of information tasks and subject domain to support enhanced discovery and access to folktales. In: Eleventh International ISKO Conference - Paradigms and conceptual systems in knowledge organization, 2010, Rome, Italy, February 23-26. *Proceedings of the Eleventh International ISKO Conference*. Würzburg: Ergon, 2010.

TILLEY, C. L.; LA BARRE, K. A. What if they build it and no one comes? Balancing Full-Text Access and User Tasks. In: Twelfth International ISKO Conference - Categories, contexts and relations in knowledge organization, 2012, Mysoure, India, August 6-9. *Proceedings of the Twelfth International ISKO Conference*. Würzburg: Ergon, 2012.

WILLIAMSON, N. J. Knowledge structures and the Internet: Progress and Prospects. In: Ninth International ISKO Conference - Knowledge organization for a global learning societ, 2006, Vienna, Austria, July 4-7. *Proceedings of the Ninth International ISKO Conference*. Würzburg: Ergon, 2006.

WILSON, T. D. Human Information Behavior. *Information Science Research*, v. 3, n. 2, 2000.

Subject representation in Archival Science: an analysis of national and international scientific articles



Leolibia Luana Linden

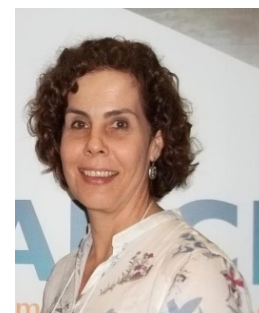
Federal University of Santa Catarina
leolibialuana@gmail.com

Graziela Martins de Medeiros
Federal University of Santa Catarina
graziela.m@ufsc.br



Luciane Paula Vital

Federal University of Santa Catarina
luciane.vital@ufsc.br



Marisa Brascher

Federal University of Santa Catarina
marisa.brascher@ufsc.br

Introduction

The term archival information is recent in the literature and derives from a paradigm shift, in which the treatment moves from the document to the information represented in it. This shift brings Archival Science even closer to Information Science, as it proposes to study the phenomena related to information, and, consequently, issues related to access, highlighting the descriptive and thematic treatment.

In this treatment, subject representation is worth emphasizing, as it requires a theoretical deepening in the scope of Archival Science, considering the studies developed in Information Science, without disregarding the specificities of the archival document, such as the fact of maintaining relations between them, that need to be explained through archival principles. It is fundamental to consider the principles that govern Archival Science to prevent the mischaracterization of documents.

Thus, this study aimed to verify how subject representation in archival documents is treated in the publications of the area of Information Science. For that, we searched the national and international databases, verifying the different aspects presented by the authors. This work constitutes a first exploratory approximation of the thematic, without the pretension to carry out a deep discussion on the identified theoretical approaches.

Organization and representation of archival documents

The organization and representation of information in archives are developed specifically in the classification and description processes. The classification refers to the physical and intellectual ordering of documents, in a hierarchical way and with explicit relations between them. The archival description refers to the information organization and representation aiming at managing and retrieving the documents. The definition of archival description presented in the General International Standard Archival Description (ISAD (G)), allows to affirm that it constitutes a process of information organization and representation:

Elaborating an accurate representation of a description unit and its component parts, if any, by extracting, analyzing, organizing and recording information that serves to identify, manage, locate and explain archival documents and the context and archival system that produced them. (INTERNATIONAL COUNCIL ON ARCHIVES, 2000, p.4).

In Brazil, the description is backed by an international standard, already cited, ISAD (G), and a national standard, the Brazilian Standard for Archival Description (NOBRADE), published in 2006. In relation to the international standard, Nobrade adds the area of access points and subjects indexing, in relation to the international guideline and, in this aspect, it seems to us to advance in the consolidation of the understanding of archival description as an activity of Information Organization and Representation of (IOR). However, according to Oliveira (2012, p.64) "It is clear that the area dialogues with the issue of access points in an inconsistent way when the elements extrapolate the more traditionally admitted [...]". The traditional elements pointed out by the author are, for example, producer, date and series. In this case, the author does not mention authorship, publication data and the title itself, which goes against what the norm considers - that the documentary items may contain the authorship. A terminological imprecision is observed in this issue.

Ribeiro (1996, p.114), referring to the Portuguese reality, and after research in archival institutions, concludes that there is a need for access points by subject, and states that: "we also need both nationally and internationally, guiding principles on indexing, which are easily applicable to the reality of the Archives."

In this perspective, this paper presents a research on the papers that has discussed subject representation in archival documents, with the objective of contributing with the discussion on this subject.

Methodology

The research is characterized as qualitative and exploratory, considering that the data analysis aims to identify articles that address conceptual and applied aspects of subject representation in archival documents.

We performed advanced searches in LISA, BRAPCI and BENANCIB databases, in the title, abstract and keywords fields. We used the following terms in Brazilian Portuguese: representação de assunto (subject representation), indexação (indexing), análise de assunto (subject analysis) and representação da informação (information representation), combined with the term (arquivologia) Archival Science. No temporal cut was performed.

In total, 17 articles addressing the theme were identified and analyzed: 10 in LISA, 4 in BRAPCI and 3 in BENANCIB. It is noteworthy that we excluded the documents we did not have access to the full text. The analyzed articles were classified into two categories:

Category 1 - Conceptual articles: Documents that present conceptual aspects related to subject representation (and / or subject analysis and indexing) in archival documents and / or in the context of archives.

Category 2 - Articles with practical application and research in progress: Documents that present methodological applications and ongoing research concerning the process of subject representation in archival documents.

Chart 1 presents the references, in alphabetical order, identified and analyzed in this research.

CAPONE, Vera Lucia Punzi Barcelos; CORDEIRO, Rosa Inês de Novais. O arquivo fotográfico ilustrativo dos trabalhos geográficos de campo do Instituto Brasileiro de Geografia e Estatística (IBGE): proposta de uma matriz para análise documentária da paisagem. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 14., 2013, Florianópolis. **Anais...** Florianópolis: UFSC, 2013.

CZECK, R L H. Archival MARC records and finding aids in the context of end-user subject access to archival collections. **American Archivist**, v.61, p. 426-440, out., 1998.

DOOLEY, J. M. Subject indexing in context. **American Archivist**, Estados Unidos, v. 55, n. 2, p. 344-354, Apr. 1992.

ESTEBAN NAVARRO, M. A.. La representación y la organización del conocimiento en los archivos. In: ACTAS DEL I ENCUENTRO DE ISKO-ESPAÑA. **Anais...** Madrid, 4 y 5 de noviembre de 1993. Zaragoza: Universidad, Librería General. 1995. p. 65-90.

GAGNON-ARGUIN, L. Documentary analysis 2: Current problems relating to documentary analysis. **Archives (Quebec)**, Quebec, v. 28, n. 3, p. 23-41, 1997.

DELGADO GOMEZ, Alejandro. Metadatos y descripción como representación del contexto archivístico. **Ibersid: Revista de sistemas de informacion y documentacion**, Zaragoza, Espanha, p. 157-164, 2007.

HUDON, M. Indexing and documentary languages in archival environments in the era of new information technologies. **Archives (Quebec)**, Quebec, v. 29, n. 1, p. 75-98, 1998.

LARADE, S P; PELLETIER, J M. Mediating in a neutral environment: gender-inclusive or neutral language in archival description. **Archivaria**, v.35,abr., p. 99-109, 1993.

LEVESQUE, Michel L'indexation: luxe ou necessite? **Archives (Quebec)**, Quebec, v. 33, n.1, p. 17-45, 2001.

LOPEZ, André Porto Ancona; CARVALHO, Pedro Davi Silva. A classificação arquivística por assunto em documentos fotográficos: o exemplo do arquivo público do Distrito Federal. **Perspectivas em Gestão & Conhecimento**, v. 3, n. 2, p. 271-279, 2013

SANTOS, Maria José Veloso da Costa. Correspondência científica de Bertha Lutz: um estudo de aplicação da lei de Zipf e ponto de transição de Goffman em um arquivo pessoal. **Ponto de Acesso**, v. 3, n. 3, p. 317-326, 2009.

SCHENKOLEWSKI-KROLL, Silvia; TRACTINSKY, Assaf. Archival description, information retrieval, and the construction of thesauri in Israeli archives. **Archival Science**, v. 6, n.1, p. 69-107, mar. 2006,.

SIMÕES, Maria da Graça; FREITAS, M. Cristina V. de. A classificação em arquivos e em bibliotecas à luz da teoria da classificação: pontos de convergência e de divergência. **Ponto de Acesso**, Salvador, v. 7, n. 1, p. 81-115, 2013.

SILVA, Márcio Bezerra da; SOUSA, Dulce Elizabeth Lima de; BANDEIRA, Pablo Matias. A representação temática em documentos arquivísticos: o caso da indexação documental realizada pelos alunos de arquivologia da UFPB. **InCID: Revista de Ciência da Informação e Documentação**, São Paulo, v. 3, n. 2, 2012.

SOUSA, Renato Tarciso Barbosa de. A representação da informação: classificação e indexação automática de documentos de arquivo. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 15., 2014, Belo Horizonte. **Anais...** Belo Horizonte: UFMG, 2014.

YAKEL, Elizabeth. Archival representation. **Archival Science**, v.3, n.1, p. 1-25, 2003.

Chart 1 – Analyzed references

Analysis and discussion of results

Among the 17 analyzed articles, 11 were classified in category 1, which we discuss below.

Of the 17 selected articles, 11 were considered conceptual. The archival classification and description are presented as processes of information and knowledge organization and representation by Simões and Freitas (2013), Esteban Navarro (1995) and Yakel (2003).

On the other hand, Yakel (2003) defines the term 'archival representation' as the most assertive today to name the archivist's job of (re)ordering, interpreting, creating substitutes and 'structures' for representation systems.

In the process of organizing and representing archival documents, Lopez and Carvalho (2013) emphasize the need to respect the principles of this area, especially provenience and organicity. In this sense, Capone and Cordeiro (2013), for example, present a proposal of analysis of photographs of the rural landscape establishing categories that include context information. This interpretation is shared in the articles that were part of the analysis, both among national and international authors, such as: Gagnon-Arguin (1997), Hudon (1998), Lopez and Carvalho (2013), Capone and Cordeiro (2013), Sousa (2014). To consider the archival principles, for the aforementioned authors, is to ensure historical reconstitution and, thus, the evidential value of the documentation.

Other authors add the use of documentary languages to this discussion. Schenkolewski-kroll and Tractinsky (2006) discuss the need for a standardization in the use of documentary languages, especially the thesaurus, in the indexing process of archival documents. They stress the need to have a minimum structure, adapted to the reality of each archive, such as the archival description standards. The authors present two thesauri used in Israeli archives, in the context of information retrieval in automated systems, that is, focused on the technological issue. Gagnon-Arguin (1997) argues that the use of controlled vocabulary and thesaurus has been discussed among Quebec archivists and points out that this is a challenge for the area, as well as Dooley (1992), who verifies the need for subject access in archives in a more consistent way, also considering the characteristics of archival documents (material form, time, geographical places, functions).

With another approach, the language was also pointed out by Larade and Pelletier (1993) as a fundamental element in the process of representing documents in libraries and archives. They emphasize that language is partial, fruit of political positions and interferes in information access and retrieval. It questions the way in which archival institutions position themselves in the process of language bias in information representation.

Three articles published in Canada have been retrieved: Gagnon-Arguin (1997), Hudon (1998) and Lévesque (2001), who are convergent regarding the point of view

of subject representation, since they relate it to the description process and consider the subject as an additional form of access.

Gagnon-Arguin (1997) points out that the treatment of subject in archives has been a growing concern among Québec archivists since 1986, expressed through publications on the issue and by initiatives to use subject indexing or subject headings related to the archival description rules, as a way to improve access to archival documents. In this publication, we observe strong concern with the users and their forms of search in the information retrieval process.

Hudon (1998) directly relates the process of indexing subjects with the archival description and levels of description, stating that any indexing policy will necessarily be linked to the definition of current description policies. The author adds that information organization and structuring by content is only an additional access. She indicates the effective establishment of indexing policies, indexing systems and documentary languages by archivists as challenges.

For Lévesque (2001), archivists focus on the description, minimizing the importance of indexing. That is, archivists tend to favor other forms of access to archival documents other than their content. By showing classification and description exercises, followed by theoretical discussion, it shows that the context is the most privileged point of view. However, in the author's view, the archivist should help the researcher or user obtain different forms of access to the documents.

It seems to us that the Canadian authors, specifically Quebecers, are pioneers on the theme, since they point to the existence of publications and institutional initiatives in the country from 1986. They are convergent when they point out the access to the document by the provenance as the main one, in the scope of the archival science, and point to a gap in studies on subject representation.

Regarding category 2, related to the practical applications and / or research in progress, we classified a total of 6 articles, about which we now address.

Santos (2009) presents a practical application of Zipf's law and Goffman's Point of Transition in the process of automatic content representation in archival documents, in the context of a personal archive. However, the author does not discuss the relevance of subject representation process in archives and the specificities of that environment.

Also addressing automatic indexing, Silva, Sousa and Bandeira (2012) present a practical application of indexing and creation of indexes with archival documents,

based on the guidelines of Noprade's area 8. They emphasize that the thematic information representation in Archival Science is in the process of technical and practical development. In this technology application thematic, Czeck (1998) analyzes the content of research tools (guides, indexes, inventories) of archival collections, aiming to determine how the MARC record can influence the representation of this information.

In the category ongoing research, the articles identified, published in Brazil, were Aguiar and Kobashi (2013) and Sousa (2014). In Spain, Delgado Gomez (2007) addresses information representation as a way of maintaining, during the documentary lifecycle, its value of proof. It analyzes and discusses metadata models, such as ISO 23081 - Metadata for document management and the Encoded Archival Description (EAD), a standard for representation of research tools in Web environments, proposing a future multidimensional application.

Aguiar and Kobashi (2013) highlight the possibility of dialogue between information organization and representation, from the area of Information Science, and Archival Science, in an area that could be named "Organization and Representation of Archival Information (ORAI). This work aims to synthesize the main theoretical-conceptual and methodological contributions from Knowledge Organization and Representation as a possibility in the processes and systems of Archival Organization and Representation.

Sousa (2014) presents a research that aims at the construction of a model for automatic classification and indexing of archival documents, preserving the principles of respect for funds and provenance. The author discusses what would be the subject of archival records and makes notes that other aspects of archival material can provide important clues to the subject of the document, strengthening the relationship of provenance to content without emphasis on one or the other. The analyzed papers show us that the problematic is being approached today, in the search of elements that contribute to methodologies of archival documents representation.

Conclusions

As evidenced by the results, the organization and representation of archival information has been discussed with greater emphasis since the 1990s. Canada has a focus on theoretical work, which discusses subject representation related to archival principles.

The searches made in the databases that were part of the research allowed to verify that the number of documents that actually refer to the proposed theme are very small. We verify that the articles selected are, for the most part, exploratory, as opposed to the more in-depth conceptual discussions. This can be considered a relevant data for the area, since it denotes an expressed need to carry out theoretical-conceptual research.

As for the documents that address conceptual aspects, we observe a concern with the aspects related to the relation between the representation of subjects and the archival functions of classification and description. We also observe the emphatic defense for considering the principles of organicity and provenance in any process of organization and representation of archival documents.

The papers with practical applications and ongoing research show a tendency to highlight technology, such as automatic representation of content and metadata standards.

Thus, it can be concluded that the topic is little explored in articles retrieved in the databases of the area of Information Science, although the authors emphasize the need to deepen the discussions and, even, to propose studies in this direction.

References

The references was made following ABNT rules.

BORKO, H. Information science: what is it? *American Documentation*, v. 19, n. 1, p. 3-5, jan. 1968.

BRASIL. Conselho Nacional de Arquivos. *NOBRADE*: Norma Brasileira de Descrição Arquivística. Rio de Janeiro: Arquivo Nacional, 2006. Disponível em: <http://www.conarq.arquivonacional.gov.br/Media/publicacoes/nobrade.pdf>. Acesso em: 04 jul. 2012.

CONSELHO INTERNACIONAL DE ARQUIVOS. *ISAD(G)*: norma internacional de descrição arquivística. 2.ed. Rio de Janeiro: Arquivo Nacional, 2000. (Publicações técnicas, n. 49). Disponível em: http://www.conarq.arquivonacional.gov.br/Media/publicacoes/isad_g_2001.pdf. Acesso em: 04 jul. 2012.

HEREDIA HERRERA. A. *Archivística General: teoría y práctica*. 5.ed. Sevilla : Diputación de Sevilla, 1991.

RIBEIRO, F. *Indexação e controle de autoridades em arquivo*. Porto: Câmara Municipal do Porto, Arquivo Histórico, 1996.

_____. Organizar e representar a informação: apenas um meio para viabilizar o acesso? *Ciência e Técnicas do Patrimônio*: Revista da Faculdade de Letras, Porto, v. 4, p. 1-22, 2005.

SILVA, I.O.S. *A organização e a representação do conhecimento no domínio da arquivística*. 2012, 193f. Tese (Doutorado em Ciência da Informação) - Universidade Estadual Paulista, Marília, 2012.

OLIVEIRA, A. L.V. *Descrição e pesquisa: reflexões em torno dos arquivos pessoais*. Rio de Janeiro: Móbile Editorial, 2012.

Analysis of methodology in bibliometric studies: a proposal for context indicators



Maria Guiomar da Cunha Frota
Federal University of Minas Gerais
guiomar@eci.ufmg.br



Ana Cláudia Ribeiro
Federal University of Minas Gerais
anacribeiro29@gmail.com

Introduction

Scientific production is responsible for the consolidation of new knowledge, result of the process and development that occur through research. It plays an important role in the scientific and technological development of countries.

Bibliometrics allows to analyze and evaluate the processes related to scientific activity. Thus, bibliometric studies have been established over the years as an important method of analysis and one of the most appropriate techniques of study for the visualization of scientific production. According to Guedes and Borschiver (2005), Bibliometrics is a set of laws and empirical principles that contribute to establish the theoretical foundations of Information Science.

The analysis of knowledge production allows us to create reliable methods to show society how a particular area has been developing, its trends and a community's behavior. For Macias-Chapula (1998), Bibliometrics has been shown as a tool capable of assisting in the observation of the state of science and technology through the production of scientific literature, at a certain level of specialization.

The creation of indicators for publications allows the scientific community and researchers to reflect on the evolution of the area and the path they are taking in what is related to a certain subject or discipline. This helps the research activity itself, since the visualization of the area allows monitoring the study unfolding, cooperating with the scientific development.

It is intended, more broadly, to investigate the methodology used in studies referring to the book HAYASHI, Maria Cristina P. Innocentini; MUGNAINI, Rogério.;

HAYASHI, Carlos Roberto Massao (Orgs); Encontro Brasileiro de Bibliometria e Cientometria. **Bibliometria e cientometria: metodologias e aplicações**. São Carlos: Pedro & João, 2013. And test the methodology in the journal **Sur Revista Internacional de Direitos Humanos**.

Specifically, the methodological steps for data observation were as follows:

- Retrieve scientific articles from the book to use in successive stages of the work.
- Identify the most used indicators in recent bibliometric studies according to the literature
- Analyze these indicators, having as balancers, the relevant characteristics of each one.
- Reflect on the pertinence of creating new indicators capable of ascertaining the socio-political insertion of a given publication.

The *Sur Revista Internacional de Direitos Humanos* was created in 2002 to bring together scholars from the southern hemisphere who work in the field of human rights and to promote their cooperation with UN agencies. The Journal still ranks according to the latest classification of Capes in A2 and according to the database, the only one ranked Qualis A, focusing on Human Rights.

The South American countries have been developing considerably in social and economic levels in recent years, so the importance of discussing, reflecting and expanding the role of human rights as a central element of human relations must be contemplated by governmental means in public policies and in universities through debates, research, publications and training courses for professionals in various fields.

Theoretical references

Human Rights

Article 19 of the Universal Declaration of Human Rights, 1946 states: "Everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers "(UN, 2009). Freedom of information is recognized as a fundamental right and the cornerstone of all freedoms. Access to information is an individual right related to the right to freedom of expression and autonomy.

Information Science has an indispensable role in what refers to Article 19 of the Universal Declaration of Human Rights. Configuring as an interdisciplinary and socially responsible field of knowledge, through its performance or scientific production, created to meet the information needs of society, therefore, a social science that interrelates with other areas of knowledge, concerned with meeting the information expectations and needs of society.

Considering the breadth and importance of Human Rights in terms of scientific literature diversity, it is reasonable to think that the theme is becoming increasingly complex and heterogeneous. These factors indicate the social and cultural relevance that the field has been gaining over the years.

In this sense, and under the scope of Information Science, knowledge must be organized, so that the advancement produced by researchers is transformed into information accessible to the scientific community (Macias-Chapula, 1998). Scientific activity has a cumulative character, based on consensus. This accumulated knowledge represents the collective memory of science, and materializes in the form of technical-scientific literature. When analyzed through information measurement techniques, they can point to trends in the development of society, in scientific disciplines, as well as in the areas of production and consumption (WORMELL, 1998).

Despite research on Human Rights has intensified after World War II, there is still a lack of studies on the scientific production of this area.

Thus, it is opportune to carry out a diagnosis on the bibliometric methodology indicators applicable to this area and on new indicators, which may enhance the analysis of the socio-political context. The importance of creating new indicators will be carried out from a bibliometric analysis of the *Revista Sur de Direitos Humanos*.

It should be emphasized that studies of this nature are of fundamental importance both for the field of Human Rights and for Information Science, since they affirm the social and interdisciplinary character, as well as their vocation for the study of properties, behavior and forces governing the flow of scientific information.

Sur Revista Internacional De Direitos Humanos

The *Sur Revista Internacional de Direitos Humanos* was created in 2002 to bring together scholars from the southern hemisphere who work in the field of human rights and to promote their cooperation with UN agencies.

The Journal has been strengthening ties between scholars and human rights defenders from Global South. *Revista Sur* aims to broaden the voices and participation of these actors within international organizations and universities. The main focus of the journal is the Southern Cone countries and their policies related to Human Rights.

Bibliometrics

According to Price (1976) studying the scientific production allows knowing and characterizing the tendencies and gaps of science. The importance of scientific evaluation came from the invention of the press in the fifteenth century, when there was a great increase in the availability of printed texts in Europe, which triggered the growth of book production. This growth in information sources has had an impact on information dissemination, as well as on the evaluation of what was disclosed. However, the publications increased exponentially, but an evaluation system was lacking.

According to Macias-Chapula (1998), Bibliometrics allows the study of the quantitative aspects of the production, dissemination and use of registered information and develops mathematical models and models to measure these processes, using the results to elaborate provisions and support decision making.

To carry out the bibliometric study, it is necessary to know which parameters relate better to the research. Such parameters are known as "bibliometric laws". These laws are defined by Vanti (2002, p.153) according to the following chart:

Chart 1 – Bibliometric laws

Lotka's law or inverse-square law	Zipf's law or Principle of least effort	Bradford's law or Bradford's law of scattering
Points to the mediation of authors' productivity, using a distribution model of the frequency size of the various authors in a set of documents.	Consists of measuring the frequency of the appearance of words in various texts, generating an ordered list of terms of a particular discipline or subject.	By measuring the productivity of journals, it is possible to establish the nucleus and the areas of dispersion on a certain subject in the same set of journals.

Thus, Bibliometrics has been used in various areas of knowledge as a methodology for developing indicators in the evaluation of scientific production.

Methodologies associated to software also indicate the paths to be followed by the scientific community and contribute to information exchange, as well as the transfer and management of knowledge in a selected field of study.

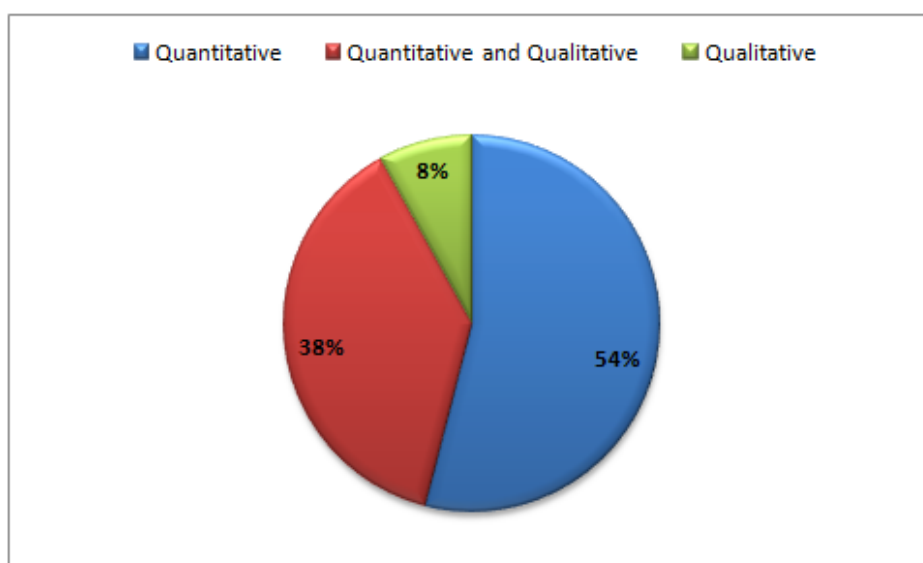
Methodological Procedures

The adopted methodology is exploratory and descriptive in two stages: 1) a bibliometric approach to the survey of the main indicators included in the book: HAYASHI, Maria Cristina P. Innocentini; MUGNAINI, Rogério.; HAYASHI, Carlos Roberto Massao (*Orgs*); Encontro Brasileiro de Bibliometria e Cientometria. **Bibliometria e cientometria: metodologias e aplicações**. São Carlos: Pedro & João, 2013., and 2) the application of bibliometrics to the articles of Sur Revista de Direitos Humanos (2011-2013).

The data were systematized with the purpose of analyzing first the indicators of the articles in bibliometrics, in order to analyze these indicators. A total of 13 articles published in the book *Bibliometria e Cientiometria: metodologias e aplicações* were analyzed, afterwards 50 articles of Sur Revista Internacional de Direitos Humanos (2011-2013) were analyzed.

Analysis of the indicators in bibliometric studies

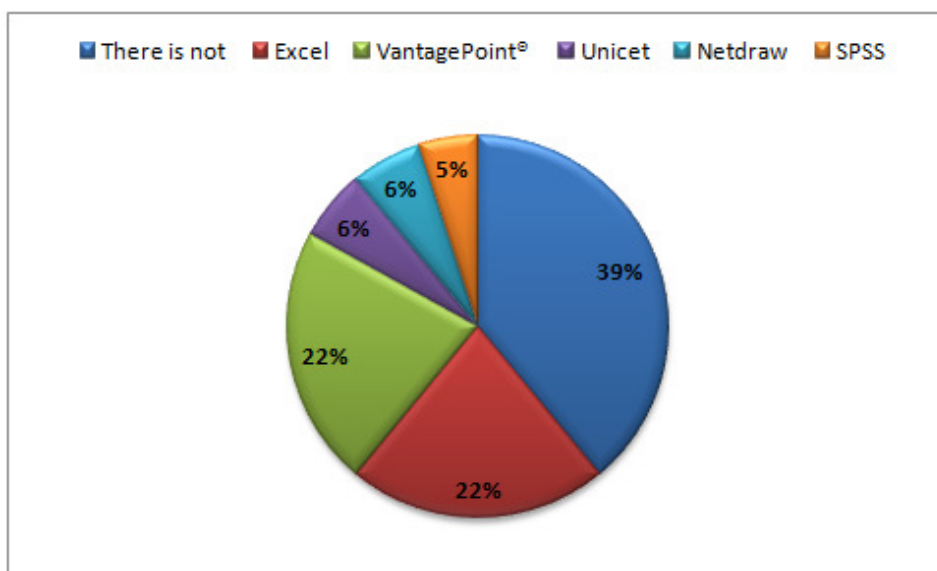
Graph 1 – Research approach



Source: by the authors

Graph 1 represents the highest concentration of articles with a quantitative approach of 54%, followed by the qualitative approach 38% and quantitative and qualitative approach 8%.

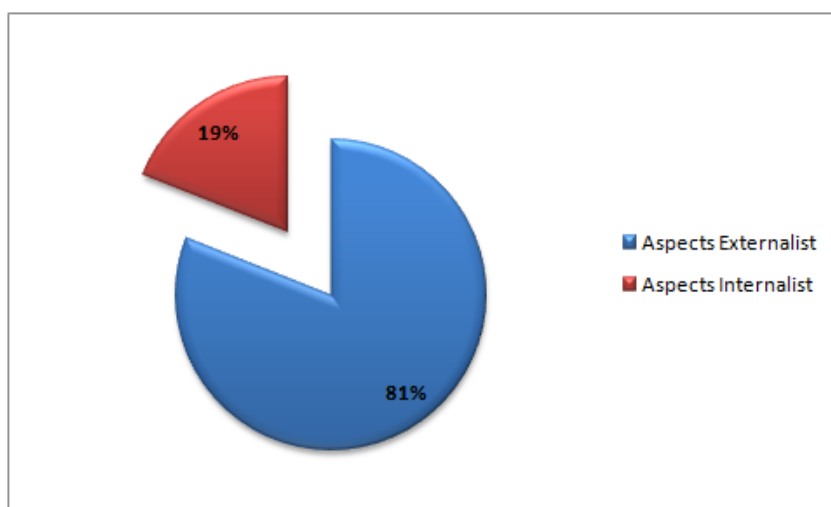
Graph 2 – Software



Source: by the authors

In Graph 2, most studies did not use any software resulting in 39%; Excel resulted in 22%, however, it is not considered a bibliometric software; the most used bibliometric software was VantagePoint® with 22%; followed by Unicet and Netdraw each with 6% and SPSS with 5%. It is noted that using software is of paramount importance for data processing and composition and research.

Graph 3 – Nature of bibliometric studies



Source: by the authors

The external aspects aim to quantify data, such as:

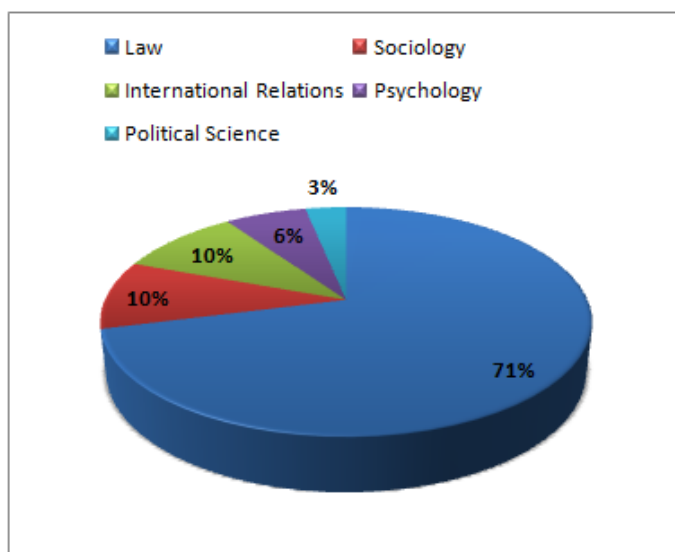
1. the temporal and geographical distribution of scientific production (when and where studies were carried out);
2. the researchers' behavior and the participation of research centers (who were the authors and evaluators, studies, and in which institutions were carried out);
3. the represented areas of knowledge (areas to which the studies are related).

As for the internal aspects, they aim to characterize the content of the scientific papers, quantifying data as the terms used by researchers to describe studies in the areas of knowledge to which these researches are related and the keywords that indicate the topics addressed by the research.

Authors' Titles in Sur Revista Internacional de Direitos Humanos (2011-2013)

Analyzed indicators: Undergraduate, Masters and Doctorate of the authors. From this perspective of the authors' analysis, it is possible to construct an indicator of the socio-political context. As the Human Rights production is heterogeneous and interdisciplinary, it will be possible to compose the areas of knowledge that have been contributing most to human rights issues.

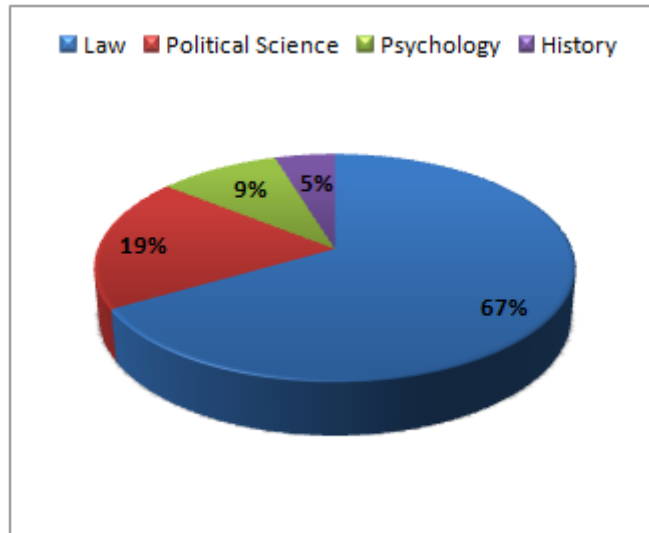
Graph 4 – Authors' titles



Source: by the authors

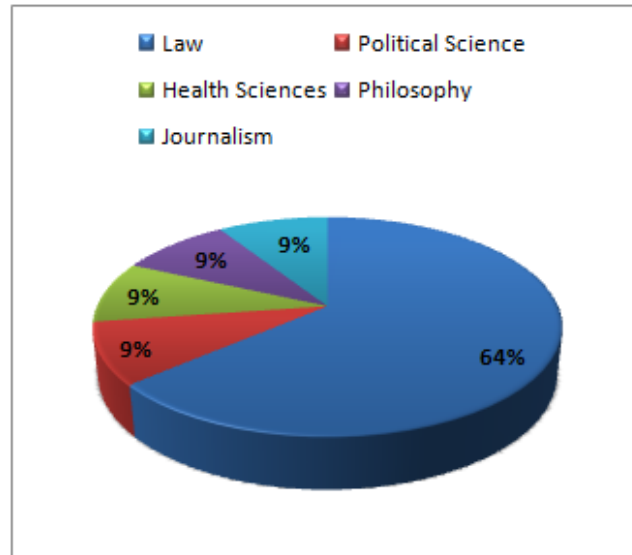
Of the analyzed 50 articles, 71% of the authors have a degree in Law, followed by degrees in International Relations and Sociology, 10% each, Psychology 6%, and Political Science 3%.

Graph 5 – Authors' Master's Degree



Source: by the authors

Graph 5 – PhD Authors



Source: by the authors

In formation in graduate courses, out of the analyzed 50 articles, 67% of the authors have master's degrees in Law, followed by Political Science 19%; Psychology 9% and History 5%.

In PhD formation, 50 articles were analyzed, 64% in Law, followed by Political Science, Health Sciences and Philosophy, each with 9%. These indicators are

relevant, since the authors' formation demonstrates the areas of knowledge that have been presenting the greatest contributions to the area of Human Rights.

This research aims, not only to quantify and verify, but also to assign meaning to the data, qualifying them so that they can have better use in scientific production in Human Rights, and that the research reverberates for society, either by each specialty or research group, or in broader regional, national or global contexts. According to Santos and Kobashi (2009), "[...] metric studies are increasingly approaching the so-called soft sciences because the latter offer theories and models that allow interpreting data in different cultural, political, Ideological and economic contexts. "

The creation of new indicators for publications that allow the scientific community, and the researchers themselves, to reflect on the evolution of the area and the path they are taking in what is related to a certain subject or discipline helps the research as the visualization of the area allows monitoring the study unfolding, cooperating with the scientific development.

Conclusions

The discussion on the theme bibliometric indicators begins with the very mission of bibliometrics, which is of allowing visualization of a certain area of knowledge. It is part of bibliometrics to follow the unfolding, to point out trends and to reflect on the object of study itself. It is necessary to study the general and specific characteristics of each area of knowledge to establish forms of visualization that contemplate the particularities of each area of scientific production.

There is a concern to test and establish new indicators, due to the research that is under development of the master's degree with the research group: Memory, repression and reparation to human rights violations in the Southern Cone countries (ECI / UFMG). And thus, testing new effective indicators that can be contemplated in the bibliometric analyzes, as there is a gap in research on scientific production in Human Rights that goes beyond quantitative analysis.

In this way, this work sought to propose a debate in order to stimulate the innovative investigation using bibliometric indicators that conform to specific aspects of the Human Rights theme, preserving the analysis of the context that surrounds this theme.

References

The references was made following the ABNT rules

BUFREM, L.; PRATES, Y. O saber científico registrado e as práticas de mensuração da informação. *Ciência da Informação*. Brasília, v. 34, n.2, p.9-25, maio/ago. 2005. Disponível em: <<http://revista.ibict.br/ciinf/index.php/ciinf/article/viewArticle/682>>. Acesso em: 08 abril de 2014.

GIL, A. C. *Como elaborar projetos de pesquisa*. 5. ed. São Paulo: Atlas, 2010.

GUEDES, V.; BORSCHIVER, S. *Bibliometria: uma ferramenta estatística para a gestão da informação e do conhecimento, em sistemas de informação, de comunicação e de avaliação científica e tecnológica*. In: CIFORM – ENCONTRO NACIONAL DE CIÊNCIA DA INFORMAÇÃO, 6., 2005, Salvador. Anais... Salvador: ICI/UFBA, 2005.

HAYASHI, Maria Cristina P. Innocentini; MUGNAINI, Rogério.; HAYASHI, Carlos Roberto Massao (*Orgs*); Encontro Brasileiro de Bibliometria e Cientometria. *Bibliometria e cientometria: metodologias e aplicações*. São Carlos: Pedro & João, 2013.

MACIAS-CHAPULA, C. A. O papel da informetria e da cienciometria e sua perspectiva nacional e internacional. *Ciência da Informação*, Brasília, v. 27, n. 2, p. 134-140, maio/ago. 1998. Disponível em: <<http://revista.ibict.br/ciinf/index.php/ciinf/article/view/342>>. Acesso em: 03 março de 2014.

OLIVEIRA, M. de. Origens e Evolução da Ciência da Informação. In: OLIVEIRA, M. de. (org). *Ciência da Informação e Biblioteconomia: novos conteúdos e espaços de atuação*. Belo Horizonte: Editora UFMG, 2005. p.9-28.

ORGANIZAÇÃO DAS NAÇÕES UNIDAS (ONU). *Direitos humanos*. Disponível em:<<http://www.un.org/en/rights/>>. Acesso em: 23 março de 2015.

ORGANIZAÇÃO DAS NAÇÕES UNIDAS (ONU). *A Declaração Universal dos Direitos Humanos*. 2009. Disponível em:<<http://www.dudh.org.br/definicao/documentos/>>. Acesso em: 23 março de 2015.

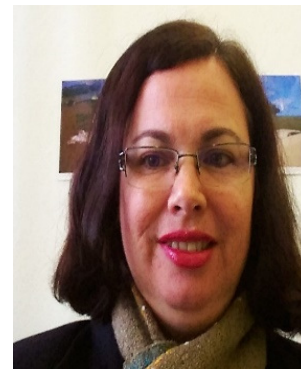
PRICE, D. de S. *O desenvolvimento da ciência: análise histórica, filosófica, sociológico e econômica*. Rio de Janeiro: Livros Técnicos e Científicos, 1976.

SUR Revista internacional de direitos humanos. São Paulo: Conectas, 2014. Disponível em: < <http://www.surjournal.org/sobresur19.php>>. Acesso em: 10 maio de 2015.

VANTI, N. A. P. Da bibliometria à webometria: uma exploração conceitual dos mecanismos utilizados para medir o registro da informação e a difusão do conhecimento. *Ciência da Informação*, Brasília, v.31, n.2, p. 152-162, maio./ago. 2002. Disponível em: < http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19652002000200016&lang=pt>. Acesso em: 15 maio de 2015.

WORMELL, I. Informetria: explorando bases de dados como instrumentos de análise. *Ciência da Informação*, v. 27, n. 2, p. 210-216. 1998. Disponível em: < <http://revista.ibict.br/ciinf/index.php/ciinf/article/viewArticle/367>>. Acesso em: 25 março de 2015.

Knowledge Organization and Representation: contribution to metric studies



Cibele Araújo Marques dos Santos
University of São Paulo
cibeleac@usp.br

Introduction

This study aims to identify the contribution of Knowledge Organization and Representation (KOR) to metric studies of the scientific production in Information Science in Brazil which allows to develop methodologies for thematic indicators of research.

The scientific production grows in unlimited proportion in the current society due to technological resources and the multiplication of tools to disseminate this information, including the access to information allowed by the web. Likewise, significant challenges are observed regarding the development of research policies, especially in universities, which are major producers of scientific and technological information in Brazil.

In this context, the analysis of scientific production becomes a fundamental resource for planning national and institutional scientific policies, and bibliometric indicators can identify patterns and trends in the areas. These metrics allow measures of activity, impact, collaboration and dynamics of the results of scientific communication represented by indicators, analyzing the data of publications and citations in a quantitative way (SANTIN, BRAMBILLA, STUMPF, 2013).

On the other hand, the analysis of the topic of scientific production allows a more qualitative view as it maps the themes and research coverage, and the existing gaps can be verified. However, the approaches to this type of analysis generate debate regarding the most appropriate methods to represent a disciplinary domain based on

a thematic structuring, regarding the levels of subjectivity, validity and updating of the classification model used (LIBERATORE and HERRERO SANTANA, 2013). The dispersion resulting from the variability and granularity of subjects addressed in a set of journal articles is significant when we analyze the keywords of each article, in addition to inconsistencies, use of synonyms, linguistic variations, as well as differentiated criteria of leveling the information treatment, especially in interdisciplinary areas. Knowledge Organization and Representation tools can help overcome these difficulties in order to carry out a more consistent mapping.

Methodology

An exploratory research was carried out in Brazilian journal articles to identify the methodology used in the definition of the themes in Information Sciences in metric studies.

Bibliographical research was carried out in Brapci database with a search strategy, carried out in May 2015, using the field keywords, searching the terms: metric studies (resulted in 9 articles), Bibliometrics (120 articles), bibliometric analysis (18 articles), thematic analysis (2 articles, only one in bibliometrics) and mapping (15 articles); the total results brought some articles that were not pertinent to the thematic bibliometrics.

The articles were selected from the title and abstract, aiming to analyze the resources of Knowledge Organization and Representation used by the authors to carry out thematic mappings.

Metric studies: characteristics and thematic

Several articles, papers in proceedings, dissertations and theses with quantitative and qualitative bibliometric approaches analyze the scientific production in Information Science from indicators related to authors, such as affiliation and co-authorship, references and author citations, journals and co-citations, but not many papers include thematic analysis.

The scientific production, as pointed out by Liberatore et al (2007), considering teachers and researchers at graduate level, is analyzed by sources, methodologies, themes, research lines and forms of dissemination. The authors, in a consistent review papers, report that the theses defended in the programs were also studied in the issues related to their development, issues, objects of study and modes of citation.

In this sense, Macedo (1987) published an article on research in Information Science and Librarianship that addressed, among other issues, the theme theses, and the data were collected in the Catalogs of Dissertation and Theses in Information Science and Librarianship published between 1982 and 1985 by IBICT. The subjects identified for theses and dissertations were taken from indexing made by the bibliographic source of these data, the cited catalogs, and used the thesaurus constructed from the document "Percepção estrutural da Ciência da Informação e áreas correlatas" ("Structural Perception of Information Science and related areas") with the following divisions: Generalities; Teaching, research and professional activity; Organization of information activities and libraries; Literature and document studies; Entry, treatment, information storage; Information retrieval and dissemination, information transference and use and related areas. In the mentioned article, the theses were distributed by major subjects with title and geographic data and also analyzed by the more specific subjects of the thesaurus.

This division of subjects was also used according to Instituto Brasileiro de Informação em Ciência e Tecnologia (1984) by Bibliografia Brasileira de Ciência da Informação that references articles from journals in the 1980s.

Poblacion and Noronha carried out several studies on the production of graduate courses in Brazil from Núcleo de Produção Científica at ECA-USP. In an article published in 2002, they analyzed the production by characterizing it thematically through the lines of institutionalized research of its authors in the graduate programs of five Brazilian universities and included 22 different lines of research that contained common themes, but dispersed the subjects of the documents.

In another paper, Queiroz and Noronha (2004), thematically mapped the dissertations and theses of ECA-USP graduate program using the thematic classification Broad Subject Headings, the subject heading list of the Library and Information Science Abstracts (LISA), which, in its printed version, presents 19 thematic categories subdivided into specific subjects. The authors report that each of the 114 items was classified in only one of the 19 categories on the list, plus a subcategory within this main subject, in addition to one the program's research lines. The categories used were: Librarianship and Information Science; Professions; Libraries and information centers; Use of libraries and users; Materials; Organization; Library buildings; Technology in libraries; Technical services; Information dissemination; Bibliographic control; Bibliographic record; Information storage and

retrieval; Information and Communication Technology; Reading; Media; Knowledge and learning; Records management; other related subjects.

Mueller and Pecegueiro (2001), analyzed articles from the journal *Ciência da Informação* from 1990-1999, which considered as significant elements of the scientific production of the area, indicating the predominant themes and comparing their frequency and the descriptors registered in LISA database, which presented expressive differences. A table with 9 macro-themes was used: Information Science, Librarianship and Documentation; Teaching, professional activity and research; Organization and management of information activities, libraries and research centers; and user studies, information transference and use and library; Literature and document studies; Library buildings; Technical services; Entry, information processing, storage, retrieval and dissemination, other related or additional subjects.

Percegueiro (2002) published another paper on the theme of six Brazilian journals in Information Science that were classified by the previous table now with the addition of the theme "Information policy and scientific and technological policy". The classification was compared again with the occurrence of subjects in the LISA base and differences were also found. In this paper, the author draws attention to the variation of certain themes over the years, for example the peak of publication on Teaching, professional activity and research in 1990 falling to 5 articles in 1991 and 2 in 1992. The theme Buildings did not have any publication in the period and Technical Services had few published articles.

Knowledge Organization and Representation and Metrics

The contribution of KO to metric studies, in the studies highlighted above, shows classification procedures by macro subjects or by specific subjects, the use of IBICT's thesaurus, the list of LISA terms database and controlled vocabulary adapted for thematic metric studies. The use of these documentary languages occurred through indexing and sometimes re-indexing of the data collected, but in most studies the linguistic and terminological issues involved in the choice of vocabularies and the methodologies of documentary analysis procedures were not discussed.

In a study about the journal *Ciência da Informação*, Silva (2002), quickly discusses aspects related to controlled vocabulary and language issues in the indexes for thematic analysis of scientific production. Vanz and Stumpf (2010), discuss production evaluation processes including sources of data collection, national and

international databases, Google Scholar and others, used software, procedures for cleaning and standardization of data considered essential to eliminate inconsistencies, as well as the organization of production in areas of knowledge. The authors consider that the insertion of production in large areas becomes a way of avoiding the overlapping of subjects, such as the use of classification schemes such as the 15 ISI knowledge divisions or the CNPq-CAPES tables of knowledge areas.

In the research carried out for the bibliometric application in the analysis of dissertations and theses, Kobashi and Santos (2010), propose an interdisciplinary approach of Social Studies of Science, Knowledge Organization and Representation and advanced bibliometric methods. The authors point out that bibliometric studies are traditionally based on description and quantification and propose to combine the approach of quantitative and qualitative studies. In these studies, relationship indicators are constructed by co-occurrence, be it authorship, citations or words, and can be applied in the elaboration of structured maps of knowledge and relationship networks, considering that these maps are representations of the scientific production of the analyzed area. There is also the possibility of using concepts produced by the area itself, which brings "literary warrant in the generation of cartographic representations of the cognitive institutionalization of a given field of knowledge."

The use of authors' keywords in the thematic analysis was recurrent, and Araujo and Melo (2011) in the analysis of the articles published in the journal *Perspectivas em Ciência da Informação*, opted to use them and observed that the lack of terminological control of keywords causes an "immense variation in the way of expressing certain concepts of the area". They commented that, when possible, they performed terminological standardization and that the dispersion of the terms and the different categories used as keywords prevented a more in-depth analysis, allowing only the visualization of a suggestive panorama of the addressed topics.

In turn, Liberatore and Herrero-Solana (2013) analyzed the themes of scientific investigation in Information Science between 2000 and 2009 studying articles from 4 journals in the area, and in their methodological approach they chose not to adopt a specific scheme using the authors' keywords because they understood it to be a representation adjusted to the perception of the authors to their field of semantic performance. They obtained a list of 2498 words belonging to 965 articles, analyzed this universe by calculating co-occurrence, setting a minimum of 8 occurrences as a significant representation.

They also carried out an analysis of the articles by macro themes of ANCIB working groups. The data matrix resulting from these analyzes together was presented through a social network whose nodes indicated by the number of relations between the occurrences allowed to identify the degree and the centrality of these relations. In the results, the authors draw attention to the components of the subjectivity of the authors in the themes, the question of the level of consensus in the terminology used by the scientific community and the dispersion of the expressions used.

Conclusions

A non-exhaustive bibliographical research was performed, selecting articles that had carried out thematic mapping in bibliometric studies, analyzing Knowledge Organization and Representation and organization procedures employed in the methodology of these works. It was observed that the authors usually do not refer to the indexing process in detail with the methodology and criteria, for example, showing only the list of terms used, if it belongs to a thesaurus or institutional vocabulary, not discussing the present categories.

There was no predominant list of subjects, but we found articles that discussed issues about language and the use of large thematic areas in procedures, the importance of an interdisciplinary approach, and the use of KO resources.

However, the articles did not present discussions about the influence of the terminologies used in the research results, although in one study a terminological standardization occurred, and in another study an article that worked with authors' keywords, the aspect of subjectivity regarding the involved choice was emphasized.

We can observe that the contribution of KO to metric studies is not exhaustive in the raised points nor in the identified papers, and that there is still a field of joint research that can be more explored allowing the improvement of the methodologies and the deepening of thematic mappings in Information Science.

References

The references was made following the ABNT rules.

ARAÚJO, C. A. A.; MELO, M. O. T. Análise dos quinze anos do periódico *Perspectivas em Ciência da Informação*. *Perspectivas em Ciência da Informação*, v.16, n.4, p.243-256, 2011.

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA. *Bibliografia Brasileira de Ciência da Informação* v.6 (1980;83). Brasília, IBICT, 1984.

KOBASHI, N. Y.; SANTOS, R. N. M. dos. Arqueologia do trabalho imaterial: uma aplicação bibliométrica a análise de dissertações e teses. *Encontros Bibli*, n. esp. p. 106-115, 2008.

LIBERATORE, G.; HERRERO-SOLANA, V. Caracterización temática de la investigación em Ciencia de la Información en Brasil en el período 2000-2009. *Transinformação*, v.25, n.3, p.225-235, 2013.

LIBERATORE, G.; HERRERO-SOLANA, V.; GUIMARÃES, J. A. C. Análise bibliométrica do periódico brasileiro Ciência da Informação durante o período 2000-2004. *Brazilian Journal of Information Science*, v.1, n.2, p. 3-21, 2007.

MACEDO, N. D. de. Pesquisa em Ciência da Informação e Biblioteconomia: implicações na pós-graduação; análise temática. *Ciência da Informação*, v.16, n.2, p.129-144, 1987.

MUELLER, S. P. M.; PECEGUEIRO, C. M. P. de. O periódico Ciência da Informação na década de 90: um retrato da área refletido em seus artigos. *Ciência da Informação*, v. 30, n. 2, p.47-63, 2001.

PECEGUEIRO, C. M. P. de. Temática dos artigos de periódicos brasileiros na área da ciência da informação na década de 90. *Transinformação*, v.14, n.2, p.117-131, 2002.

POBLACION, D. A.; NORONHA, D. P. Produção das literaturas “brancas” e “cinzenta” pelos docentes-doutores dos programas de pós-graduação em ciência da informação no Brasil. *Ciência da Informação*, v.31, n.2, p.98-106, 2002.

QUEIROZ, F. M.; NORONHA, D. P. Temática das dissertações e teses em Ciência da Informação no Programa de pós-graduação em Ciências da Comunicação da USP. *Ciência da Informação*, v.33, n.2, p. 132-142, 2004.

SILVA, F. M. e. Análise da revista Ciência da Informação disponibilizada na SciELO a partir do seu vocabulário controlado. *Transinformação*, v.14, n.2, p.133-138, 2002.

VANZ, S. A. de S.; STUMPF, I. R. C. Procedimentos e ferramentas aplicados aos estudos bibliométricos. *Informação & Sociedade*, v.20, n.2, p.67-75.

Funding: FAPESP Project number 2013-04109-7

Relativizing indexes h and g: a study applied to the domain of metric studies



Deise Deolindo Silva
São Paulo State University
deisedeolindo@hotmail.com



Maria Cláudia Cabrini Grácio
São Paulo State University
cabrini@marilia.unesp.br

Introduction

By reflecting the development degree of a nation and by providing guidelines in the social, political and economic spheres, the evaluation of the scientific production of a country has become an indispensable factor, as well as the elaboration of criteria that evaluate the various scientific domains.

Among the approaches to the characterization and evaluation of a domain, bibliometric studies, especially citation-based analyzes, stand out as constituting a consolidated and objective approach (HJØRLAND, 2002).

In this context, emphasis is placed on efforts to develop adequate mechanisms for evaluating scientific domains, as well as to provide tools and indicators to define guidelines, allocate investments and resources, formulate programs and evaluate activities related to the scientific and technological development as a way of identifying the scientific elites, guaranteeing a profitable investment in research by the sponsoring agencies. (MUGNAINI *et al.*, 2004, THOMAZ *et al.*, 2011, VANZ; STUMPF, 2010).

Given this need, Hirsch (2005) developed a bibliometric index, named h-index, for evaluating the scientific performance of researchers; h-index simultaneously considers aspects related to the impact (citations) and productivity (number of articles) of the researcher.

This index presents some characteristics that should be considered during the analysis of the researchers' performance: it should not be used to compare

researchers from different areas and may attribute advantages to those who have a longer title career.

Egghe (2010) notes that new researchers have a clear disadvantage when compared to those with a longer career. In addition, the h-index may increase, even if there are no new published articles.

This author has developed a variation of h named g-index, "we define the g-index as the highest rank such that the first g papers have at least g^2 citations together" (EGGHE, 2010, p. 13). One of the characteristics of this indicator is to give greater weight to the number of citations received by the article. However, h index can also assign higher values to researchers with longer publication periods.

In this context, this study considers that the relativization of these indicators according to the researcher's publication time can eliminate the dependence of the h index and g index in relation to the time (t), allowing a more reliable visualization of the impact and insertion of the researcher in the scientific community.

For the h-index, the relative h/t indicator is defined by: given the author's h index and the publication time t, we obtain the h/t ratio, according to which the author published, on average, h/t articles with at least h/t citations each, per year.

For the g index, the relative g/t indicator is defined by: given the author's g index and the publication time t, we obtain the g/t ratio, according to which the researcher published, on average, g/t articles per year, which together have at least g^2/t citations.

Considering the above, this research aims to perform a comparative analysis between h-index, g-index and their relativized form by publication time, of the most productive researchers in "Metrics Studies", thematic assigned to the area of Information Science, in order to evaluate trends and adequacy of the indicators in the description of the impact and insertion of the researchers in the scientific community.

More specifically, we seek to identify, describe and analyze the h-index, g-index and the relative indexes h/t and g/t, taking as the universe the most productive researchers in the 1980-2014 period, from the journal *Scientometrics*, considered representative of the subject Metrics Studies in mainstream science.

This research is justified by the contribution to the deepening of the analysis of h type indexes by Hirsch and the relationship between them. It will help understanding these indicators as descriptors of the academic trajectory of a researcher in the subject Metric Studies.

Methodological procedures

In order to verify the behavior of the indicators, the 20 most productive authors in the journal *Scientometrics* were taken as the universe, corresponding to having at least 20 scientific articles published during the 35-year period (1980-2014). From February 6 to February 10, 2015, we collected in Scopus database, the articles published by each researcher, the number of citations per article and the year of publication.

From the data collected, we calculated the h and g indexes, the ratios h/t and g/t of each researcher. For each indicator, we obtained: mean; standard deviation; maximum; minimum; coefficient of variation and Pearson's correlations, in order to verify the tendency and scatter of the group regarding the studied indicators and the correlation between the absolute h and g indices and the relative h/t and g/t.

Analysis and interpretation of results

Table 1 presents the values of h and g indexes and their relativizations by publication time of the 20 most productive researchers in the journal *Scientometrics*, ordered according to the h index of the researchers.

It is verified that the author with the highest h-index is Glänzel ($h = 36$). The author has been publishing in *Scientometrics* for 31 years. In contrast, Gupta has the lowest index ($h = 5$) and has published in *Scientometrics* for 12 years, a relatively short period when compared to the first researcher.

The average h-index of researchers is around 15, meaning they have at least 15 articles with at least 15 citations each. In addition, it can be said that the scatter of the group is 7.4 articles with at least 7.4 citations each.

It is noteworthy to verify the behavior of h-index relativized by the period of time (h/t) the researchers have been publishing in the subject "Metric Studies" in the *Scientometrics Magazine*.

Glänzel has published in *Scientometrics* for 31 years, his h-index is 36 and the h/t ratio is 1.16, that is, he has published, on average, 1.16 articles with at least 1.16 citations each, per year. Guan and Thelwall have a h/t ratio of 1, they have published one article with at least one citation, per year.

Table 1. Bibliometric indicators for the 20 most productive authors in Scientometrics (1980-2014)

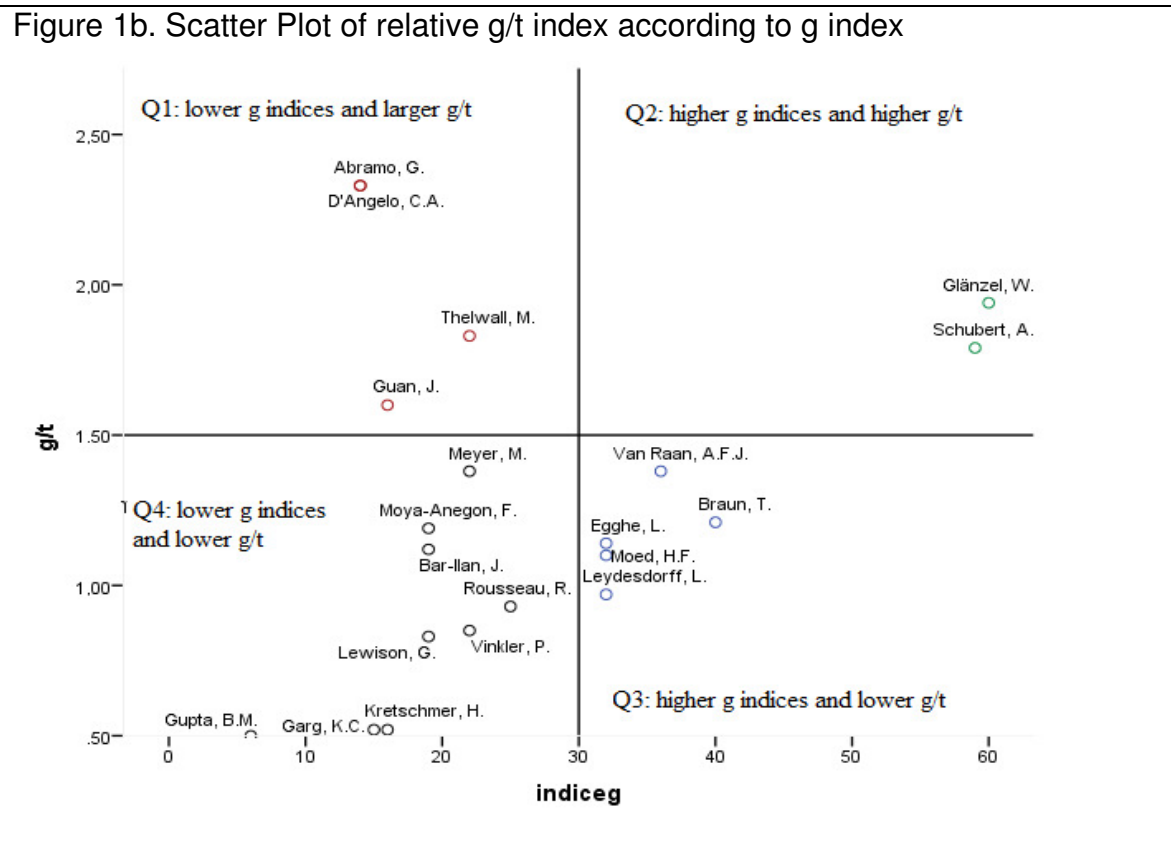
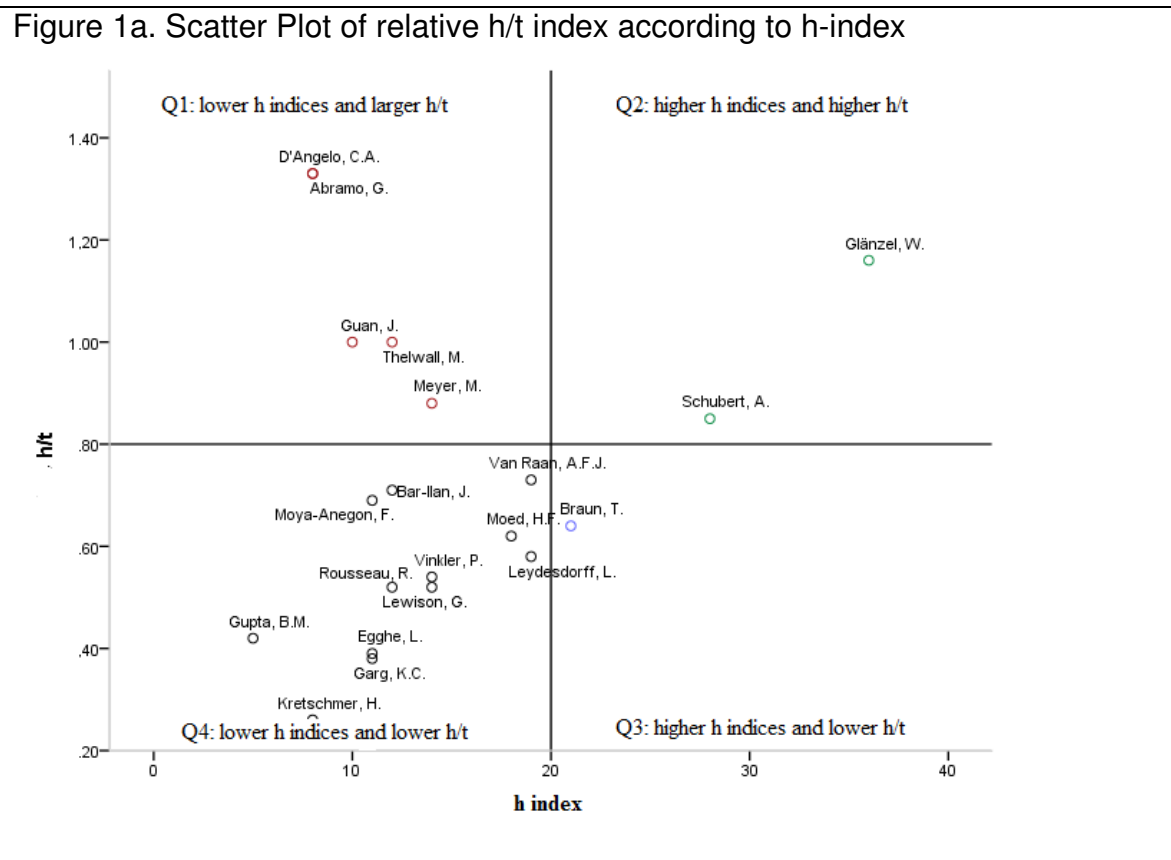
Researcher	H-Index	G-Index	Time t (years)	Ratio h/t	Ratio g/t
Glänzel, W.	36	60	31	1.16	1.94
Schubert, A.	28	59	33	0.85	1.79
Braun, T.	21	40	33	0.64	1.21
Leydesdorff, L.	19	32	33	0.58	0.97
Van Raan, A.F.J.	19	36	26	0.73	1.38
Moed, H.F.	18	32	29	0.62	1.10
Meyer, M.	14	22	16	0.88	1.38
Rousseau, R.	14	25	27	0.52	0.93
Vinkler, P.	14	22	26	0.54	0.85
Lewison, G.	12	19	23	0.52	0.83
Thelwall, M.	12	22	12	1.00	1.83
Bar-Ilan, J.	12	19	17	0.71	1.12
Moya-Anegon, F.	11	19	16	0.69	1.19
Egghe, L.	11	32	28	0.39	1.14
Garg, K.C.	11	15	29	0.38	0.52
Guan, J.	10	16	10	1.00	1.60
Abramo, G.	8	14	6	1.33	2.33
D'Angelo, C.A.	8	14	6	1.33	2.33
Kretschmer, H.	8	16	31	0.26	0.52
Gupta, B.M.	5	6	12	0.42	0.50
Mean	15.0	26.0	22,2	0.73	1.27
Standard Deviation	7.4	14.2	9,3	0.31	0.55
Maximum	36	60	33	1.33	2.33
Minimum	5	6	6	0.26	0.50
Coefficient of Variation	51%	55%	42%	42%	43%

Source: by the authors

Researchers Abramo and D'Angelo published 6 years ago, and have an h-index of 8 and $h/t = 1.33$, that is, they published, on average, 1.33 articles with at least 1.33 citations each year. It may be noted that the authors published in this journal a few years ago, however, these articles were of relevance and were reflected in the h-index.

Figures 1a and 1b allow a better visualization of the scatter of the relative indexes according to the absolute h and g indices. It is observed, in Figure 1a, that out of the 20 analyzed researchers, five (25%) of the researchers presented quite different trends between the index and their derivatives, relativized by publication time:

D'Angelo, Abramo, Guan, Thelwall and Meyer presented lower h indexes and high h/t indexes, on the other hand Braun presented high h-index and lower value for h/t.



Source: by the authors

In addition, we observe authors who have published for many years in the journal and present relatively low index, such as Kretschmer, Garg, and Egghe. Most of the analyzed authors publish in *Scientometrics*, in the studied subject, on average, one article every two years.

When considering the relative h/t index, the most recent authors match the oldest ones in publication time, by eliminating the dependence on publication time. This relative indicator overcomes the disadvantage presented by the h index, in which researchers with longer titration time have greater advantages when compared to more recent researchers.

Pearson's correlation between these two indexes equal to 0.203 shows that the trend of the values of these two indicators does not present a statistically significant association, as they represent different characteristics of the researchers.

In relation to g-index, it identifies similarly that Glänzel has the highest g-index and Gupta the lowest ($g = 6$). The average g-index of authors is 26, meaning: this is the largest number, in which the most cited papers received together, at least 676 citations (26^2 citations). It should be noted that the variability of the group is practically twice as that observed for the h-index.

The g/t ratio analyzes the quotient between g-index and publication time in the journal. On average, researchers have a quotient of 1.27, this is the highest number the most cited papers received together, at least 1.6 citations ($1,27^2$ citations) per year.

Figure 1b shows that Abramo, D'Angelo, Thelwall and Guan have high values for relativized indices (g / t) and lower values of g-index, evidencing different trends in relation to these two indexes.

As for the h and h/t indexes, Pearson's correlation between g and g/t, equal to 0.288, showing that there is no statistically significant association between them.

When analyzing the relative h/t and g/t indicators, almost the same authors with high productivity and impact are identified, thus, these metrics also help the metric analysis of researchers in a theme.

Conclusions

It can be verified that h and g indexes portray well the scientific performance of researchers in the subject Metrics Studies. It is observed that indicator h does not value highly cited articles nor those with few citations. On the other hand, g-index places

more weight on the number of citations the article received. Thus, it is considered that the indicators can identify different aspects of the researcher's impact and are complementary in a possible evaluation.

H/t and g/t ratios can complement an author's evaluation when researchers with different publication times are analyzed, since this quotient eliminates the influence of time. Thus, the first relation determines the h-index per year and the second determines the citations received jointly by published articles, per year.

Finally, it is considered that in order to reliably evaluate the scientific performance of a researcher, theme, discipline, area of knowledge or country it is indicated to use a combination of these indicators in the performed analysis.

References

The references was made following the ABNT rules.

EGGHE, L. The Hirsch index and related impact measures. *Annual review of information science and technology*, White Plains, v. 44, p. 65-114, 2010.

HIRSCH, J. E. *An index to quantify an individual's scientific research output*, 2005.

HJØRLAND, B. Domain analysis in information science: eleven approaches traditional as well as innovative. *Journal of Documentation*, v. 58, n. 4, p.422-462, 2002

MUGNAINI, R.; JANNUZZI, P. M.; QUONIAM, L. *Indicadores bibliométricos da produção científica brasileira: uma análise a partir da base Pascal*. *Revista Ciência e Informação*. v. 33, p. 123-131, 2004.

THOMAZ, P. G; ASSAD, R. S.; MOREIRA, L. F. P. *Uso do fator de impacto e do índice h para avaliar pesquisadores e publicações*. *Arquivo Brasileiro de Cardiologia*, p. 90-93, 2011.

VANZ, S. A. S.; STUMPF, I. R. C. *Procedimentos e ferramentas aplicados aos estudos bibliométricos*. *Informação e Sociedade*, v. 20, n.2, p. 67-75, maio/ago. 2010.

Domain Analysis in knowledge organization and representation in Archival Science



Cynthia Maria Kiyonaga Suenaga
São Paulo State University
cmksuenaga@gmail.com



Brígida Maria Nogueira Cervantes
Londrina State University
brigidacervantes@gmail.com

Introduction

This study is concerned with an analysis of documentary information aiming at knowledge, and thus seeks methodologies to structure and order stored knowledge in the form of individual documents or documentary sets in Archives.

The perspective raised by the scientific-informational paradigm (aimed at the informational content recorded in documents), in detriment to the documentary paradigm (concerned with custody and preservation), considers the document no longer as a physical object to be stored and preserved only, but rather used and analyzed in several ways, initiating a new way of thinking of archives.

Archival Science is a discipline of human knowledge that requires research work to direct its applications, a discipline of continuous reflection on the models of interpretation and intervention in reality (SOUSA, 2012). Domain Analysis, proposed by Hjørland and Albrechtsen (1995), as a new approach to the study of the field of Documentation and Information Science (IS), is a theoretical-methodological formulation which, through the union of base theories and applied procedures from IS, includes Knowledge Organization and other areas in it.

The Document Classification Plan (DCP), on the informational content of the documents and the functions and activities that gave rise to them, can contribute to the identification and understanding of the knowledge structures in the archive. According to Héon (1995), DCP is the first instrument of intellectual tracking for documents at any level. Thus, based on one of the prerogatives for the elaboration of the DCP, which is

a careful evaluation of the institution that generates or maintains the documents, the DCP is selected as the research object to be analyzed in the search to identify the existing knowledge in the Domain of an archive.

From the examination of one of the main products of documentary treatment in archives, the question is: how can Domain Analysis in the context of Knowledge Organization in IS contribute to the understanding of the knowledge structure present in Brazilian State Public Archives?

The objective of this research was to investigate knowledge structures present in Brazilian State Public Archives (APEBs), from the perspective of Domain Analysis. With this aim, we tried to identify the existence of Document Classification Plans in Brazilian State Public Archives; to verify the presence of standards for the preparation of the Document Classification Plan; to relate the Document Classification Plan to Domain and to the discursive community to select the concepts and possible relationships in the Document Classification Plan.

Theoretical Assumptions

An essential characteristic of information-as-knowledge (BUCKLAND, 1991), a perspective we adopt in this research, is its intangible form. Therefore, in order to express it, it is necessary to describe it or represent it and at the moment it becomes information-as-thing, something that can be manipulated in some way. Under this approach, knowledge representations form subsets of information-as-thing, which can materialize as documents or as Knowledge Organization Systems, which are important ways to cope with information-as-knowledge. Knowledge Organization Systems, a product of Knowledge Organization, record aspects of the document that address issues directly related to Domain and socio-historical-cultural aspects that are evidenced by Domain Analysis.

Information is a mediating agent in knowledge production and appears in form and content as significant structures capable of generating individual and collective knowledge (BARRETO, 1994). Such structures are constituted by conventions and agreements among individuals who coexist in groups within the limits of their domain.

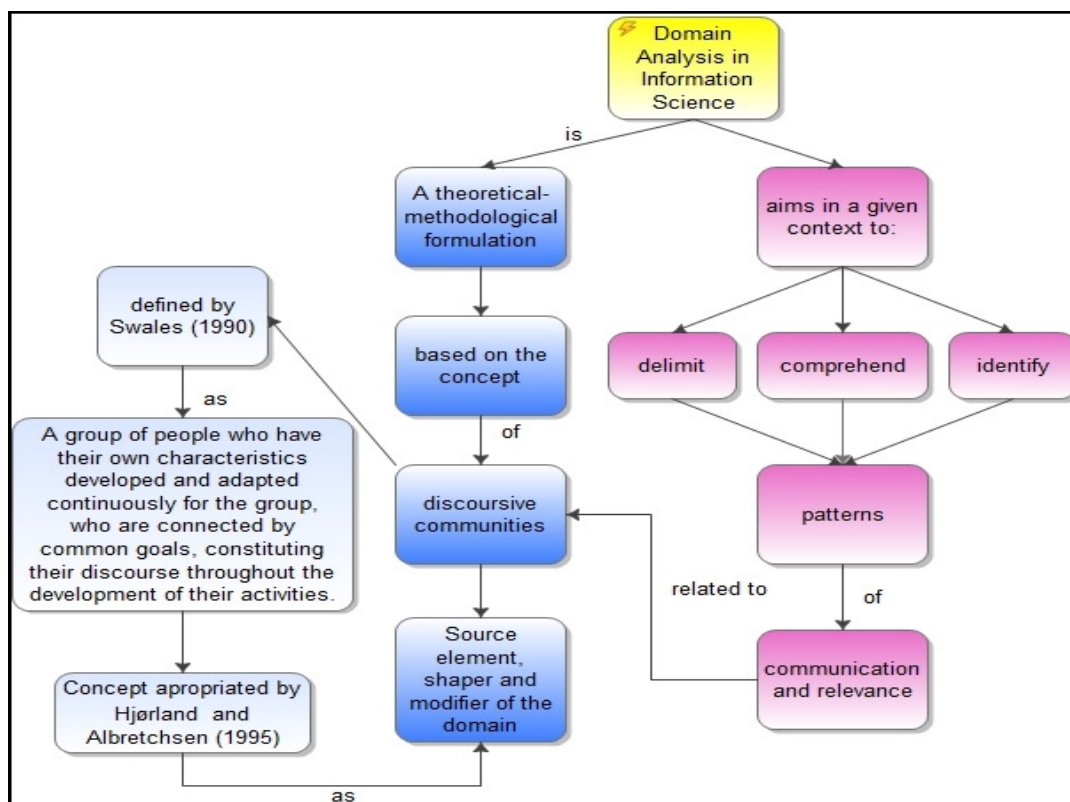
According to Hjørland and Albrechtsen (1995), such groups are the discursive communities, which play a role in society that is reflected in the knowledge organization, in the informational structure, in the forms of language and

communication, in the information systems, in the patterns and criteria of relevance. For Swales (1990), a discourse community is a group of people who connect first to pursue common goals; their communicative needs toward meeting goals or objectives tend to predominate in the development and maintenance of their characteristics.

Hjørland and Albrechtsen (1995) base the theory of Domain Analysis on the finding that the Domain is constituted by the discursive community and that the record of its communicative processes reflects the knowledge circulating in this environment. Later Hjørland (2002) presents eleven ways to approach a Domain in order to elucidate this knowledge. These approaches encompass fundamental actions of IS, such as indexing, bibliometric maps, structures of Scientific Communication, historical, epistemological, terminological and critical studies, the author states that it is advisable to use two or more approaches concomitantly and that other approaches may arise .

The following is a representation of the general lines that underlie the Domain Analysis approach in IS (Figure 1).

Figure 1 - General lines of Domain Analysis in IS.



Source: by the author

The focus of Domain Analysis is to delimit and understand the set of information in a given domain from the discourse community that constitutes and modifies it in space and time. According to Foucault (2008), discourse is not simply that which manifests or conceals desire, but also that which is the object of desire; discourse is not simply that which expresses the struggles or domination systems, but what we struggle for. The basis of this approach in the discursive community expands the research horizon of the Information Scientist, as looking at the discourse always involves historical, contextual and social variables.

Tennis (2003) seeks to bring the Domain Analysis theory closer to the actual realization of informational studies aimed at the Domain, presenting what can be named the Domain Analysis development, proposing two axes that consist of analytical devices that complement DA. According to the author, although the eleven approaches to Domain Analysis proposed by Hjørland (2002) offer subsidies to study the domain, they do not offer structured definition and systematized delimitation of what is being analyzed.

The modulation area is the axis in which, through a statement, the extension of the examined domain is named and delimited, indicating what is included or not and how the domain will be named (relevant terms are defined and the extension of the Domain). The degree of specialization qualifies and marks the intention of the Domain, decreasing its extension and increasing its intention, the domain qualifier is the focus given to the analysis (TENNIS, 2003).

According to Nordberg, Linell and Gunarson (1997), spoken events such as phone calls, meetings, conferences, among others, are intermediated, preceded and followed by written practices in the form of documents such as memoranda, reports, contracts and others. The authors consider that the effective communication in discourse is written and fundamental for society to function. This discourse occurs not only between the professional group, but also between different groups that can be considered as discursive communities.

The materiality of the discourses of the discursive community materializes through the institutionalization and socialization of the discourse, in its formalization in texts in its broad sense, including written, visual, audiovisual language. Documentary discourse materializes in documentary texts and the result of its study can be seen as a scientific knowledge constituted in the relation with the institutions and with the subjects of science, as well as the means of circulation of knowledge (SALES, 2011;

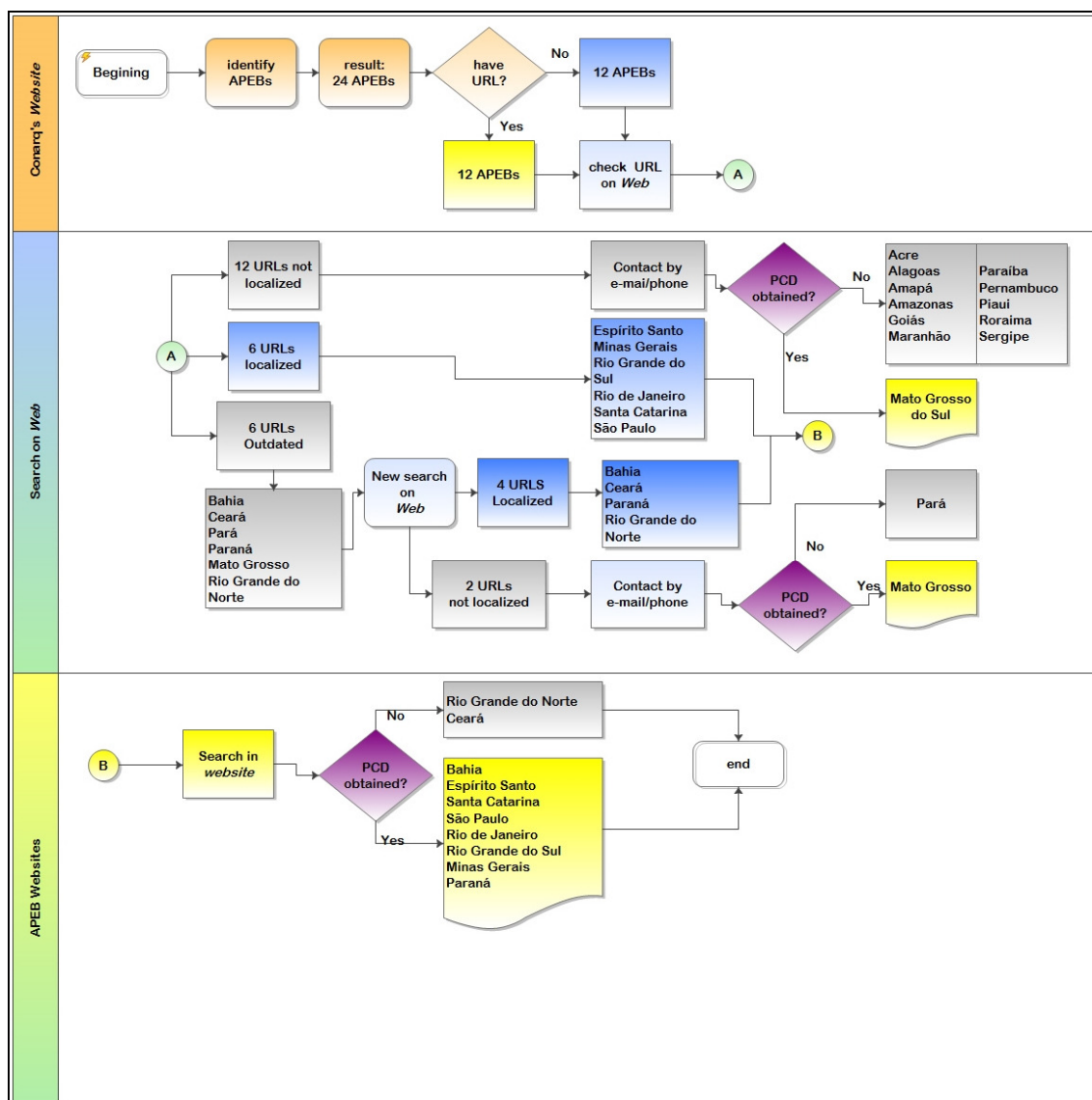
NUNES , 2008). The texts, in this direction, are materialized, institutionalized and socialized discourses through documents.

Thus, it is desirable to find ways to identify the documentary discourse to understand the formation of a discursive community and to seek to elucidate the knowledge recorded in the various domains.

Methodological procedures

A preliminary study on the Brazilian State Public Archives was carried out based on the information on Conselho Nacional de Arquivos' website (CONARQ), which provides a list of the Brazilian State Public Archives. In this list, we identified 24 State Public Archives, out of which 12 had a website address. This first selection went through a conference stage in which some URLs were found to be inaccessible. Due to these inconsistencies identified during the first access attempt, a search was made for the other APEB's websites, with no URL in the list. The following is a summary of the procedures used to bring together the Document Classification Plans.

Figure 2- Overview of the search for DCP

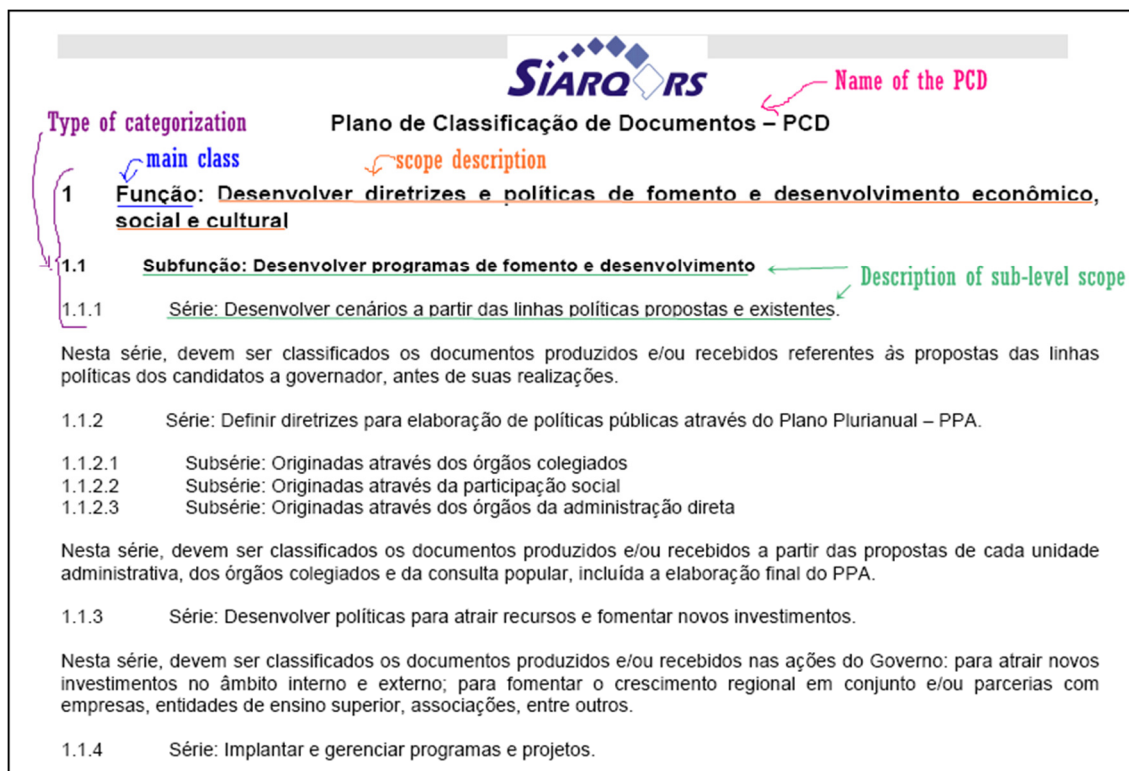


Source: by the authors

Using the 10 URLs collected from CONARQ's website and the Web, navigation was carried out in search of DCPs (Figure 2). The next step was the reading and analysis of DCP: 10 obtained PCDs, 8 per navigation 1 per e-mail and 1 per instructions via telephone.

From the reading of the plans, the most relevant information to understand the DCPs, such as their nomenclature, the description of the scope in the main classes, the description of the scope in the sub-levels, the type of categorization and the classes (Figure 3).

Figure 3 - Example of information collected in DCP



Source: by the author

This information is then analyzed in close approximation with the theorems of Tennis (2003) and Hjørland (2002) in order to support the domain-analytic approach in Archives.

Analysis and results

The APEB's DCPs, through the titles assigned to the document, somehow present information pertinent to the modulation axis by Tennis (2003), in that sense, the more specific the title, the less the extension of the Domain and the greater its intention, that is, more accurate becomes the objective or purpose of the Domain.

In the case of DCPs, it is understood that the title of the document has the function of specifying the Domain in which classification applies, that is, to clarify which documents are classified in each level in the most efficient way possible. In the example of the DCPs, most of them specify the limits in the title of the document: State of (name of the State), Executive Power, Ancillary Activities.

The reading of APEB's DCPs demonstrates that their great majority presents a description of the scope at the levels and sub-levels of categorization. It is, however,

identified that this description, where it exists, is not rigorous in all cases. Some are meticulous, such as the APEBs in Bahia and São Paulo. Others are more generic, not giving much clue as to the extent and intent of the domain.

The identification of the discourse occurred through the identification of the concepts expressed in the nomenclature presented in the entries of the DCPs. The main categories (entries), understood here by Dahlberg (1978) as concepts to their fullest extent, give indications of the extent of the sub-domain.

Chart 1 - Identification of the Discourse of APEBs main categories (as it appears in the original documents).

APE-BA	APEMS	APESP
Organização Administrativa	1 Organização Administrativa	01 Organização Administrativa
Comunicação Institucional	2 Comunicação Institucional	02 Comunicação Institucional
Gestão de Recursos Humanos	3 Gestão de Recursos Humanos	03 Gestão de Recursos Humanos
Gestão de Bens Materiais e Patrimoniais	4 Gestão de Bens Materiais e Patrimoniais	04 Gestão de Bens Materiais e Patrimoniais
Gestão Orçamentária	5 Gestão Orçamentária e Financeira	05 Gestão Orçamentária e Financeira
Gestão de Documentos e Informações	6 Gestão de Documentos e Informações	06 Gestão de Documentos e Informações
Gestão de Atividades Complementares	7 Gestão de Atividades Complementares	07 Gestão de Atividades Complementares

Source: by the authors

The similarity between the discourse of the main categories listed in the APEBs related DCPs (see Chart 1) and also the identification of the characteristics of Swales's discursive communities (1990), associating them with the APEBs (see Chart 2), indicate the constitution of a Discursive Community. The use of practically the same terms demonstrates the proximity of discourse and indicates that the goal of the discursive community is shared.

Chart 2 - Identification of APEBs as a Discursive Community.

Characteristic (Swales 1990)	Identified in the APEBs
Set of common public objectives of broad agreement	According to the law of archives: "[...] document management and special protection of archival documents, as an instrument to support administration, culture, scientific development and as evidence and information."

Mechanisms of intercommunication between its members	Use of the usual mechanisms of institutional communication, such as e-mail, offices, memorandums and others.
Use of participatory mechanisms for exchange of information and feedback	Demonstrated, for example, by the constitution of Evaluation Committees of documents and other activities that require participation of members of the team and of stern members for the joint decision making.
Desenvolvimento de conjuntos de gênero	Demonstrated by the production of Document Classification Plans, File and Document Management System, Document Disposal Term, among others.
Aquisição de léxicos específicos	Use of acronyms, such as PCD (Plano de Classificação de Documentos), TTD (Tabela de Temporalidade de Documentos), SIGAD (Sistema Integrado de Gestão Arquivística de Documentos), among others, and use of technical terms of their own (eg. atividade-meio, atividade-fim, arranjo etc).
A threshold of members with appropriate degree of relevant content and expertise in their domain.	Mandatory presence of archivists among members of a team responsible for processing records (regulated by law 6,546; 78).

Source: by the authors

The Discourse of the main categories in APEBs of Espírito Santo, Mato Grosso, Minas Gerais, Paraná, Rio de Janeiro and Santa Catarina do not present the same proximity by similarity as the APEBs of Bahia, Mato Grosso do Sul and São Paulo, but they present similarities considering the addressed concepts. The terminology used to express the categories is not the same, but the general concept expressed is the same or very similar.

Conclusions

The characterization of APEBs and consequently of other institutions such as Discursive Communities brings to researchers and professionals a theoretical-methodological base that is welcome to Knowledge Organization and especially to Archival Science. Further research aimed at the use of Domain Analysis in these fields of study is recommended, aiming to promote its development.

What is observed, in general lines, is that elements of a Discursive Community produce homogeneous discourses, although not identical. An increase in the ability to exchange information among Domain members would certainly increase the level of general expertise and new solutions to the Domain.

It is hoped that studies will be carried out to assist in the delineation of existing knowledge in Archival Science Domain, in which the Documents are object of study,

as well as their content, from the genesis of the accumulation of documents that gives origin to the Archive emphasizing the discourse, in search for the understanding of what this accumulation communicates. This type of research opens the way for Knowledge Organization to establish itself as an essential tool for Archival Science.

References

The references was made following the ABNT rules.

DAHLBERG, I. Teoria do Conceito. *Ciência da Informação*, Rio de Janeiro, v. 7, n.2, p. 101–107, 1978b. Disponível em: <<http://revista.ibict.br/ciinf/index.php/ciinf/issue/view/87/showToc>>. Acesso em: 10 maio 2013.

BARITÉ, M. Organización del conocimiento: um nuevo marco teórico-conceptual em bibliotecología y documentación. In: CARRARA, K. (Org.). *Educação Universidade e Pesquisa*. Marília: Unesp - Marília, 2001. p. 35-60.

BARRETO, A. A questão da informação. *São Paulo em Perspectiva: revista da Fundação SEADE*, v. 8, n. 4, Out-Dez, p. 3-8, 1994. Disponível em: <<http://www.seade.gov.br/produtos/spp/index.php?men=res>>. Acesso em: 20 out. 2012.

BUCKLAND, M. K. Information as thing. *Journal of American Society for Information Science*, v. 5., n. 42, p. 351-360, 1991. Disponível em: <<http://people.ischool.berkeley.edu/~buckland/thing.html>>. Acesso em: 15 jun. 2012.

FOUCAULT, M. *A ordem do discurso*. São Paulo: Loyola, 2008.

HÉON, G. Les plans de classification en archivistique. *Archives*, v. 27, n. 1, 1995. Disponível em: <http://www.archivistes.qc.ca/revuearchives/vol27_1/27-1-heon.pdf>. Acesso em: 16 fev. 2014.

HJØRLAND, B; ALBRECHTSEN, Hanne. Toward a New Horizon in Information Science: Domain-Analysis. *Journal of The American Society for Information Science: JASIS*, v.46, n.6, 400-425, 1995. Disponível em: <[http://onlinelibrary.wiley.com/doi/10.1002/\(SICI\)1097-4571\(199507\)46:>](http://onlinelibrary.wiley.com/doi/10.1002/(SICI)1097-4571(199507)46:>)>. Acesso em: 14 jul. 2012.

HJØRLAND, B. Domain analysis in information science: eleven approaches – traditional as well as innovative. *Journal of Documentation*, v.58, n.4, 422-462, 2002. Disponível em:<<http://dx.doi.org.ez78.periodicos.capes.gov.br/10.1108/00220410210431136>>. Acesso em: 16 jun. 2012.

NORDBERG, B.; LINELL P.; GUNNARSSON B. (Eds). *The construction of Professional Discourse*. London: Longman, 1997.

NUNES, J. H. O discurso documental na história das ideias linguísticas e o caso dos dicionários. *Alfa: Revista de Linguística*, São Paulo, v. 52, n. 1, p. 81-100, 2008. Disponível em: <<http://seer.fclar.unesp.br/alfa/article/view/1468/0>>. Acesso em: 28 set. 2012.

SALES, R. de. Representação de domínios em Biblioteconomia e Arquivística. In: SILVA, F. C. C. da; SALES, R. de. (Orgs.) *Cenários da Organização do Conhecimento: linguagens documentárias em cena*. Brasília: Thesaurus, 2011.

SOUSA, R. T. B. de. A classificação como função matricial do que fazer arquivístico. In: SANTOS, V. B. dos; INNARELLI, H. C.; SOUSA, R. T. (Orgs.). *Arquivística: temas contemporâneos: classificação, preservação digital, gestão do conhecimento*. 38 ed: Distrito Federal: SENAC, 2012.

SWALES, J. M. The concept of discourse community. In: SWALES, J. M *Genre analysis: English in Academic and Research Settings*. Cambridge: Cambridge University Press, 1990. p. 21–27.

TENNIS, J. T. Two Axes of Domains for Domain Analysis. *Knowledge Organization*, v. 3/4, jul. 2003. Disponível em:
<<http://search.ebscohost.com/login.aspx?direct=true&db=iih&AN=14887661&lang=pt-br&site=ehost-liv>>. Acesso em: 20 ago. 2013.

The FamilySearch Indexing as a crowdsourcing initiative in the context of knowledge organization



Paula Carina de Araújo
São Paulo State University
paula.carina.a@gmail.com



José Augusto Chaves Guimarães
São Paulo State University
guima@marilia.unesp.br

Introduction

In Information Science (IS), the importance of sharing diverse knowledge to generate new knowledge has been long discussed. This is what Wersig (1993) called the "depersonalization of knowledge" by claiming that information and communication technologies have emerged to provide the transmission of knowledge more and more, regardless of the person who possesses it. We note the development of a socio-cognitive vision, in which, not only the individual is considered, but the environment in which the subject is inserted, his/her relations and influences.

Crowdsourcing is included in this context. This term was first presented by Howe (2006) in article "The Rise of Crowdsourcing", in which the author points to examples of organizations that have recognized in mass collaboration an opportunity to access and produce cheap, relevant, and often quality content, produced by people all over the world and available on the internet.

IS inserted in postmodernity is characterized by its pluralism of methods, techniques and reflections - which leads to a greater questioning and critique of the area - and indicates that its objects begin to be taken into account, as stated by Tálamo and Smit (2007).

Hjorland (2002a), in this same line, has shown concern for strengthening the field by presenting eleven approaches to the understanding of IS and its relation to other domains. In this research, for example, the relationship between Knowledge Organization and Family History is proposed, focusing on FamilySearch Indexing

(FSI), the project has enabled the manual indexing of millions of records by hundreds of thousands of volunteers making it one of the largest crowdsourcing initiatives in the world. (HANSEN et al., 2012).

Thus, this study describes the FSI in the context of Knowledge Organization (KO). Tálamo and Smit (2007) suggest that a possible program of work in the field of information should contemplate five elements, among them, two that justify this research: - the articulation among technological devices of information, information production and the generation of meaning; - the social insertion of information, with determination of local conditions of reception, aiming at improving the devices and the study of the activity of users-consumers.

FamilySearch

FamilySearch was created in 1989 as a system that encompassed all previous automation initiatives of genealogical records collected since the creation of the Utah Genealogical Society in 1894. (Allen, Embry and Mehr, 1995).

This society is a non-profit institution that serves the community through its main library, the Family History Library located in Salt Lake City, Utah. It also has approximately 4,000 family history centers, sectorial libraries in 88 countries, and a website called FamilySearch (www.familysearch.org). Its mission is "to preserve and provide access to historical information that can be used to reconstruct ancestral and related families and to compile genealogies or family histories." (METCALFE, THATCHER, 2008, p.1).

"Prior to the Web, FamilySearch used analogue formats (floppy disks, CDs, mail, etc.) to transmit the information between headquarters and volunteers" (ELLIS, 2014, page 7). Considered an example of web collaboration, as a crowdsourcing initiative, FamilySearch has benefited greatly from the rise of the internet.

FamilySearch encourages the participation of volunteers to improve their products and services, as well as the faster availability of the records available to them. Some of these products and services can be cited: the new Family Tree; an expressive genealogical database that includes data from record indexing projects and information that comes from input from people around the world who are not necessarily members of the Church (KENNARD; KENT; BARRETT, 2011); access to genealogical records in microfilm format, to others that have already been digitized and digitized and indexed

genealogical records, which makes them retrievable through a search system and; indexing of records through the FSI, theme of this study.

The FSI was created on the initiative of The Church of Jesus Christ of Latter-Day Saints, with a view to enhancing its doctrine for the preservation of the family and the need to know the ancestors. To collaborate with the project, one needs to create an account on the FamilySearch website, download the FSI program (a web version is already being developed) on the computer and access it with the login and password. Subsequently, it is necessary to choose a batch of records compatible with its level as indexer (beginner, intermediate, advanced) and download it on the computer. These batches are sets of genealogical records, such as: birth, marriage, death, migration records, etc. (IGREJA ..., 2012).

Flinders (2012) explains that "indexing is a process of transcribing the key fields in a record or digital image. It is not a complete transcript, but a capture of the key elements of the record so it will be searchable".

Hansen et al. (2013) state that, since its inception, the FSI has been dedicated to creating high quality transcripts. To ensure correct indexing, the project has used a process to ensure quality based on a type of arbitration, in which two people (A and B) independently transcribe the information of an image (name, gender and age, for example) and any discrepancy between their work is assessed by an experienced referee (ARB) who makes the final decision. This model is called A-B-ARB.

FamilySearch Indexing (FSI) and the sociocognitive approach to knowledge organization (KO)

In studying the use of postmodern knowledge related to IS, Wersig (1993) is concerned with the changing role of knowledge for individuals, organizations and cultures, and for the author, this change involves the philosophical and technological dimension. In this perspective, the precursors of the study of the social paradigm in IS are Shera, Frohmann, Brier, Capurro and, with emphasis, Hjørland.

We propose the relationship between the FSI and the socio-cognitive approach, such as Barite (2001) considers that the KO seeks to encompass diverse practices and social activities related to access to knowledge and seeks to act as an instrument for information treatment and management of information use in order to integrate phenomena and applications related to the structuring, disposition, access and diffusion of socialized knowledge.

In the context of KO, domain analysis is presented as a theoretical approach to Information Science (IC), which "states that the best way to understand information in CI is to study the knowledge-domains as thought or discursive communities." (HJORLAND; ALBRECHTSEN, 1995, p. 400). It is understood that KO socio-cognitive approach is implicit in domain analysis.

Hjorland (2002b) explains that the central point in his approach is that tools, concepts, meanings, information structures, information needs and relevance criteria are established in discursive communities where the communication process is established. It is noticed that in this view there is a shift from the focus of Information Science, from individuals and / or computers to the cultural and scientific social world.

When Hjorland and Albrechtsen (1995) argue that it is necessary to incorporate knowledge about the cultures in which the information systems are functioning, it is possible to consider the human behavior, the relations between the participants and the indexing process in the FSI from a sociocognitive perspective in the KO domain.

This is more evident in Hjorland's (2002b, 268) statement that "an essential problem in IS is how people interpret the texts to be organized and searched as well as the information needs that should be satisfied." This approach understands individual knowledge in a historical, cultural and social perspective.

Frohmann (2008, p.20-21) argues that the concept of materiality of information "is very important when one wants to investigate what information systems do [...] in a broad sense to include what he calls information". He defends the socio-cognitive approach, which is explicit in stating that "the concept of materiality brings a much richer understanding of the public and social character of information in our time."

By understanding, just as Frohmann (2008), that information is also materialized by technological means, it is possible to affirm that the FSI can be observed and studied under this approach. One can seek the understanding of how the digital information, made available to the volunteers in the FSI, is materialized and what kind of relationships and social and political effects this initiative involves.

According to Flinders (2012, p. 23), "the strength of the program is its collaborative approach ". Following this thinking, Howe (2009, p.9) states that "although crowdsourcing is mingled with the internet, its essence is not technology." Human behaviors are most important because of "the potential of the internet to weave the mass of humanity together into a thriving, infinitely powerful organism."

The domain of the FSI involves a broad discursive community that can be recognized for the purposes of this study, such as those working on the digitization of records, volunteers indexing the records, those who act as referees when verifying the quality of the performed indexing and all who use FamilySearch to search for records and create their family trees.

It is undeniable that information and communication technologies have caused many changes in society, but it is important to realize, as Barreto (2002, 73) states, that "the real changes that come from information technologies bring to the environment a new way to create knowledge and these modifications are related to its time and space".

Conclusions

The approximation between the FSI as a crowdsourcing initiative and the socio-cognitive approach of KO demonstrated the importance of studies in this domain. We may state that because FSI involves a diverse discursive community that is made up of volunteers without technical knowledge on descriptive and thematic representation of documents that will make the records accessible on the Internet.

It is the first time that this perspective is explored in the literature, therefore, some possibilities of study could be identified: to recognize the motivating factors for volunteers to be involved with indexing in the FSI, which influences them during indexing, which tools they use as support for this process and how they influence KO, what the social function of that activity is, what power relations are involved, among others. In addition, domain analysis is considered a way of generating new knowledge about the interaction of this community with the generated information. These may be contributions from KO to the FSI.

The FSI grows every day in numbers of projects, involved in the process of transcription and making millions of records available to a large community of family history research. It is believed that their main contributions to KO domain are: identification of the relationships within this discursive community and how they influence KO, recognition of historical and social influences among these volunteers, elucidation of search patterns and use of the information available, and relationship of the KO domain with a new domain, that of family history or genealogy.

References

The references was made following the ABNT rules.

ALLEN, JB.; EMBRY, JL.; MEHR, KB. *Hearts turned to the fathers: a history of the Genealogical Society of Utah, 1894-1994*. Provo, Utah: BYU Studies, 1995.

BARITÉ, M. Organización del conocimiento: un nuevo marco teórico-conceptual en Bibliotecología y Documentación. In: CARRARA, K. (org.). *Educação, Universidade e Pesquisa*. Marília: Unesp-Marília-Publicações; São Paulo: FAPESP, 2001. p. 35-60.

BARRETO, A. A.. A condição da informação. *São Paulo em Perspectiva*, v.16, n. 3, 2002. Disponível em: <http://www.scielo.br/pdf/spp/v16n3/13563.pdf> . Acesso em: 15 out. 2009.

IGREJA DE JESUS CRISTO DOS SANTOS DOS ÚLTIMOS DIAS. *Indexação do FamilySearch*, 2012. Disponível em: https://indexing.familysearch.org/newuser/nuhome.jsf?3.15.1https://familysearch.org/volunteer/indexing_Acesso em: 10 set. 2012.

ELLIS, S. A History of collaboration, a future in crowdsourcing: positive impacts of cooperation on british librarianship. *Libri: International Journal of Libraries and Information Services*, v.64, n.1, p.1-10, 2014.

FLINDERS, S. FamilySearch indexing: providing easy access to your records. ARCHIVING, 2012, Copenhagen, Denmark. *Proceedings...* Copenhagen, Denmark: Society for Imaging Science and Technology, jan. 2012. p. 23-27.

FROHMANN, B. O caráter social, material e público da informação. In: FUJITA, M.; MARTELETO, R.; LARA, M. (Orgs). *A dimensão epistemológica da ciência da informação e suas interfaces técnicas, políticas e institucionais nos processos de produção, acesso e disseminação da informação*. São Paulo: Cultura Acadêmica; Marília: Fundepe, 2008, p.19-34.

HANSEN, D. et al. Improving indexing efficiency & quality: comparing A-B-Arbitrate and peer review. FAMILY HISTORY TECHNOLOGY WORKSHOP, 2012, Salt Lake City, UT. *Proceedings...* Salt Lake City, UT: BYU, 2012. p. 1-4. Disponível em: http://fht.byu.edu/prev_workshops/workshop12/papers/4.1%20DerekHansenEtAllIndexingExperiments.pdf Acesso em: 10 set. 2012.

HANSEN, D. et al. Quality control mechanisms for *crowdsourcing*: peer review, arbitration, & expertise at FamilySearch Indexing. In: CONFERENCE ON COMPUTER SUPPORTED COOPERATIVE WORK, 2013. *Proceedings...* ACM, 2013. p. 649-660. Disponível em: <http://dl.acm.org/citation.cfm?id=2441848> Acesso em: 10 mar. 2015.

HJORLAND, B. Domain analysis in information science. *Journal of Documentation*, Londres, v. 58, n. 4, 2002a, p. 122-162.

HJORLAND, B. Epistemology and the social-cognitive perspective in information science. *Journal of The American Society for Information Science*, v.53, n.4, p. 257-270, 2002b.

HJORLAND, B.; ALBRECHTSEN, H. Toward a New Horizon in Information Science: domain-analysis. *Journal of The American Society for Information Science*, v.46, n.6, p. 400-425. 1995.

HOWE, J. The rise of *crowdsourcing*. *Wired*, n. 1406, p. 1-4, jun. 2006. Disponível em: <http://www.wired.com/wired/archive/14.06/crowds.html> Acesso em 17 out. 2012.

HOWE, J. *O poder das multidões*. Tradução de Alessandra Mussi Araujo. Rio de Janeiro: Elsevier, 2009.

KENNARD, D.; KENT, A.; BARRETT, W. Linking the past: discovering historical social networks from documents and linking to a genealogical database. In: WORKSHOP ON HISTORICAL DOCUMENT IMAGING AND PROCESSING, 1., 2011, Provo, UT. *ACM International Conference Proceeding Series...* Provo, UT: Brigham Young University, jan. 1, 2011. p. 43-50.

METCALFE, WJ.; THATCHER, MP. Serving the genealogical and historical research communities: an overview of records access and data privacy issues. WORLD LIBRARY AND INFORMATION CONGRESS: IFLA GENERAL CONFERENCE AND COUNCIL, 74., 2008, Quebec, Canadá. *IFLA Conference Proceedings...* Quebec, Canadá: IFLA, 2008. P. 1-10.

TÁLAMO, MFGM.; SMIT, JW. Ciência da Informação: pensamento informacional e integração disciplinar. *Brazilian Journal of Information Science*, v.1, n.1, p. 33-57, jan./jun. 2007. Disponível em: <http://www2.marilia.unesp.br/revistas/index.php/bjis/article/viewFile/30/29> Acesso em: 09 out. 2012.

WERSIG, G. Information Science: the study of postmodern knowledge usage. *Information Processing & Management*, v. 29, n. 2, p. 229-239, 1993. [http://dx.doi.org/10.1016/0306-4573\(93\)90006-Y](http://dx.doi.org/10.1016/0306-4573(93)90006-Y) Acesso em: 10 out. 2012.

A study on actions to make government datasets available in linked open data



Fernando de Assis Rodrigues
São Paulo State University
fernando@elleth.org



Ricardo César Gonçalves Sant'Ana
São Paulo State University
ricardosantana@marilia.unesp.br

Introduction

The principles of Linked Open Data (LOD) establish a new way of sharing datasets opened by the Internet, aiming to promote the wide distribution of structured data in languages, such as eXtensible Markup Language (XML) and in compliance with the recommendations of the Resource Description Framework (RDF) (BERNERS-LEE, 2009; BIZER; HEATH; BERNERS-LEE, 2009; HEATH, 2015; W3C, 2014, 2015).

In this scenario, government datasets play a prominent role: they represent 18.58% of the total number of existing LOD datasets and 41.54% of these government datasets have at least one relationship with ontologies or controlled vocabularies, according to the results of the mapping developed by Linking Open Data cloud diagram (SCHMACHTENBERG; BIZER; PAULHEIM, 2014a, 2014b).

However, according to Schmachtenberg, Bizer, and Paulheim (2014a), there are still characteristics in the LOD dataset structures at the moment of data retrieval that is not considered ideal nor adopted good practices, such as the absence of metadata and licenses information.

Actions to make public government data accessible are an integral part of discussions on trends in the modernization of public administration models, which seek to redistribute skills and resources among different intra-governmental and extra-governmental organizations, allowing greater institutional pluralism in public functions (MALIN, 2006; SANT'ANA; RODRIGUES, 2013).

The strengthening of transparency actions can be expanded by building information sharing environments that, among other characteristics, provide an

increase in information flows between public administration and society, thus ensuring greater visibility of the State activities (BOHMAN, 2000; MARCONDES, JARDIM, 2003). These environments become components of greater citizen participation, extending possibilities of participation beyond voting; and the State can improve the effectiveness and monitoring of the activities and results of its actions, in addition to complying with the obligation to publish government data (BRASIL, 2011; SANT'ANA; RODRIGUES, 2013).

Access to government datasets on the results of legislative votes is important in monitoring the activities of representatives, supporting the construction of analyzes, such as "[...] the identification of party clusters" and "[...] consistency of each of our representatives in the voting during their mandates" (SANT'ANA; RODRIGUES, 2013, page 58).

The objective of this paper is to explore the actions needed to provide government datasets in Linked Open Data, starting from an application of a model of recommendations for data publication "Linked Data Best Practices in Different Topical Domains", proposed by Schmachtenberg, Bizer And Paulheim (2014a), in databases available on legislative votes of the Brazilian Senate.

The research object was delimited to datasets available in Communication and Information Technology (ICT) tools of the Brazilian Senate, more precisely on the existing votes in the 'Portal e-Cidadania - Open Data', analyzed between January and March 2015.

Methodological procedures

The methodology adopted was the exploratory analysis of research object, with a qualitative approach, through the specification of the characteristics of the existing dataset (i.e. the location of the resource on the web site, information about the descriptive page and available files); and the data structures found at the time of data collection.

Those characteristics formed a set of information that served as a subsidy for the proposal of a strategy of actions necessary for restructuring this existing data, in compliance with the established recommendations and good practices of LOD datasets availability, proposed by Schmachtenberg, Bizer, and Paulheim (2014a).

Theoretical Background

Schmachtenberg, Bizer, and Paulheim (2014a, 2014b) propose a model with recommendations for data publication, with the objective of identifying the compliance to LOD concepts and good practices for data sharing by public datasets stored in various domains. These recommendations were elaborated from community practices and the results presented by the LOD dataset mapping developed by the Linking Open Data cloud diagram (JENTZSCH; CYGANIAK; BIZER, 2011).

The model is divided into nine recommendations:

Providing Provenance Information

In the process of data retrieval, it is necessary that datasets have unique identifiers to help data retrieval process by external agents, in compliance with the first principle of LOD (BIZER; HEATH; BERNERS-LEE, 2009). These unique identifiers must conform to rules established by the Uniform Resource Identifier (URI) and the RDF (JENTZSCH; CYGANIAK; BIZER, 2011; SCHMACHTENBERG; BIZER; PAULHEIM, 2014a, 2014b).

Defining links with other datasets

The dataset needs to have links to other datasets through the relationship rules established by the RDF. This procedure facilitates the automated data collection by external agents, including others datasets to which it was linked (JENTZSCH; CYGANIAK; BIZER, 2011; SCHMACHTENBERG; BIZER; PAULHEIM, 2014a, 2014b).

Use of controlled vocabularies and existing ontologies

As data is a basic element "[...] formed by a sign or finite set of signs that do not contain, intrinsically, a semantic component, but only syntactic elements" (SANTOS; SANT'ANA, 2002, np), and it is necessary to use controlled vocabularies and ontologies to extend the semantic load at the moment of data collection by external agents, such as: Dublin Core Metadata Set (DC), Friend of a Friend (FOAF), Simple Knowledge Organization System (SKOS), among others (JENTZSCH; CYGANIAK; BIZER, 2011, SCHMACHTENBERG; BIZER; PAULHEIM, 2014a, 2014b).

Definitions of terms, elements, and attributes in vocabularies and ontologies

The additional documents linked to the dataset, containing information about ontologies and controlled vocabularies, must have unique URIs for terms, elements, and attributes (SCHMACHTENBERG; BIZER; PAULHEIM, 2014a, 2014b). For example, in FOAF the definitions that describes terms like 'name' or 'birthday', it must

be accessible by unique URIs, either by using split symbols ('/'), or hash-tag (#) to differentiate access to each term (BRICKLEY; MILLER, 2014).

Linking terms between vocabularies

In case it is necessary to develop new and own vocabularies, is important that terms of this vocabulary are linked to existing vocabulary terms, such as DC, FOAF, SKOS, among others. The linking of new vocabularies with comprehensive vocabularies provides a greater repertoire of information on terms developed for the vocabulary of external agents (JENZYK; CYGANIAK; BIZER, 2011; SCHMACHTENBERG; BIZER; PAULHEIM, 2014a, 2014b).

Providing metadata

At the time of data retrieval, datasets must have metadata elements to ensure quality on data retrieval process, to identify the data source, and to ensure quality (JENZYK; CYGANIAK; BIZER, 2011). The metadata "[...] is a key factor to minimize search and retrieval problems in the various informational environments [...]" (SANTOS; ALVES, 2009) and it is recommended that: their elements be available in the root element; and the use of DC elements (JENTZSCH; CYGANIAK; BIZER, 2011; SCHMACHTENBERG; BIZER; PAULHEIM, 2014a, 2014b).

Use of license terms in metadata

The dataset metadata must contain licenses terms in its elements and attributes, such as Creative Commons, Open Data Commons Attribution License, Open Database License (ODbL), among others (JENTZSCH; CYGANIAK; BIZER, 2011; W3C, 2011).

Providing metadata about the dataset structure

In the dataset retrieval, there must be metadata containing information about its structure - made available with the data or in supplementary documents - delimiting elements, iterations, used terms and attributes (JENZYK; CYGANIAK; BIZER, 2011; SCHMACHTENBERG; BIZER; PAULHEIM, 2014a, 2014b).

Use of alternative methods for data retrieval

The most common form of providing structured datasets in the RDF is through a SPARQL Protocol and RDF Query Language Endpoint (SPARQL Endpoint) (JENTZSCH; CYGANIAK; BIZER, 2011), which enables external agents to perform structured searches in the SPARQL query language. However, it is recommended that dump files (RDF Dump), are also available explicit in the RDF/XML standard or equivalent (SEMANTICWEB.ORG, 2011).

Dataset Characteristics

Portal e-Cidadania aims to promote transparency of actions and activities of the Brazilian Senate, through access to government data (BRAZIL, 2015a). In January 2015, the website had forty-five datasets, divided into eight groups: 'Projetos e Matérias', 'Plenário', 'Parlamentares', 'Composição', 'Comissões', 'LexML', 'Legislação', and 'Processo Legislativo'.

The group 'Plenário' contains eight subdivisions: 'Diários do Senado e do Congresso'; 'Legislaturase Sessões Legislativas'; 'Matérias com prazos'; 'Pronunciamentos de senador'; 'Questões de Ordem'; 'Sessões do Plenário'; 'Tabelas de tipos relacionados a plenário', and; 'Votações nominais' - this last containing data about votes recorded in plenary and information related to sessions, bills, votes, like subjects and votes of each member (BRAZIL, 2015b).

The subdivision 'Votações nominais' consists of 11 items: 9 dump files in XML, containing data on votes grouped annually; 1 hyperlink to a web service, and; 1 hyperlink to a voting search page - this last in HyperText Markup Language (HTML) format, is not the subject of this study.

Characteristics of data structure in retrieval process

Each dump file in XML format has a unique URL, formed by the composition: the domain/primary hierarchy 'http://legis.senado.leg.br/dadosabertos/dados/'; the prefix 'ListaVotacoes' follow by year referring to the data, and; file extension '.xml'.

In dataset retrieval via web service, its present only queries grouped by daily results. For example, to perform data collection on votes in a given month, it is necessary to perform 'x' queries, where 'x' represents the count of days in month. This retrieval of datasets has a URL for each daily result, with an URL value formed by the composition: the domain/primary hierarchy 'http://legis.senado.leg.br/dadosabertos/plenario/lista/votacao/'; the year, month and day.

In both cases - when collected the dump files and retrieved data via web service - the datasets are explicit in XML language and the collections of elements, attributes and terms available are identical (Table 1).

Element	Associated to element	Type of Data	Attributes
<i>ListaVotacoes</i>	None (Root Element)	Group Element	'xmlns:xsi'and'xsi:noNamespaceSchemaLocation'
<i>Metadados</i>	ListaVotacoes	Group Element	None
<i>Votacoes</i>	ListaVotacoes	Group Element	None
<i>Versao</i>	<i>Metadados</i>	Text	None
<i>VersaoServico</i>	<i>Metadados</i>	Integer	None
<i>DescricaoDataSet</i>	<i>Metadados</i>	Text	None
<i>Votacao</i>	<i>Votacoes</i>	Group Element	None
<i>CodigoSessao</i>	<i>Votacao</i>	Integer	None
<i>SiglaCasa</i>	<i>Votacao</i>	Text	None
<i>CodigoSessaoLegislativa</i>	<i>Votacao</i>	Integer	None
<i>TipoSessao</i>	<i>Votacao</i>	Text	None
<i>NumeroSessao</i>	<i>Votacao</i>	Integer	None
<i>DataSessao</i>	<i>Votacao</i>	Text	None
<i>Horainicio</i>	<i>Votacao</i>	Text	None
<i>CodigoTramitacao</i>	<i>Votacao</i>	Integer	None
<i>CodigoSessaoVotacao</i>	<i>Votacao</i>	Integer	None
<i>SequencialSessao</i>	<i>Votacao</i>	Integer	None
<i>Secreta</i>	<i>Votacao</i>	Text	None
<i>DescricaoVotacao</i>	<i>Votacao</i>	Text	None
<i>Resultado</i>	<i>Votacao</i>	Text	None
<i>TotalVotosSim</i>	<i>Votacao</i>	Integer	None
<i>TotalVotosNao</i>	<i>Votacao</i>	Integer	None
<i>TotalVotosAbstencao</i>	<i>Votacao</i>	Integer	None
<i>CodigoMateria</i>	<i>Votacao</i>	Integer	None
<i>SiglaMateria</i>	<i>Votacao</i>	Text	None
<i>NumeroMateria</i>	<i>Votacao</i>	Integer	None
<i>AnoMateria</i>	<i>Votacao</i>	Integer	None

Element	Associated to element	Type of Data	Attributes
<i>Votos</i>	<i>Votacao</i>	Group Element	None
<i>VotoParlamentar</i>	<i>Votos</i>	Group Element	None
<i>CodigoParlamentar</i>	<i>VotoParlamentar</i>	Integer	None
<i>NomeParlamentar</i>	<i>VotoParlamentar</i>	Text	None
<i>SexoParlamentar</i>	<i>VotoParlamentar</i>	Text	None
<i>Url</i>	<i>VotoParlamentar</i>	Text	None
<i>Foto</i>	<i>VotoParlamentar</i>	Text	None
<i>Tratamento</i>	<i>VotoParlamentar</i>	Text	None
<i>Voto</i>	<i>VotoParlamentar</i>	Text	None

Source: Authors

The root element is named 'ListaVotacoes' and has two attributes and two elements. Its two attributes are responsible for binding the dataset with a supplementary document (XML Schema), containing the delimitation of available elements, content types, and attributes.

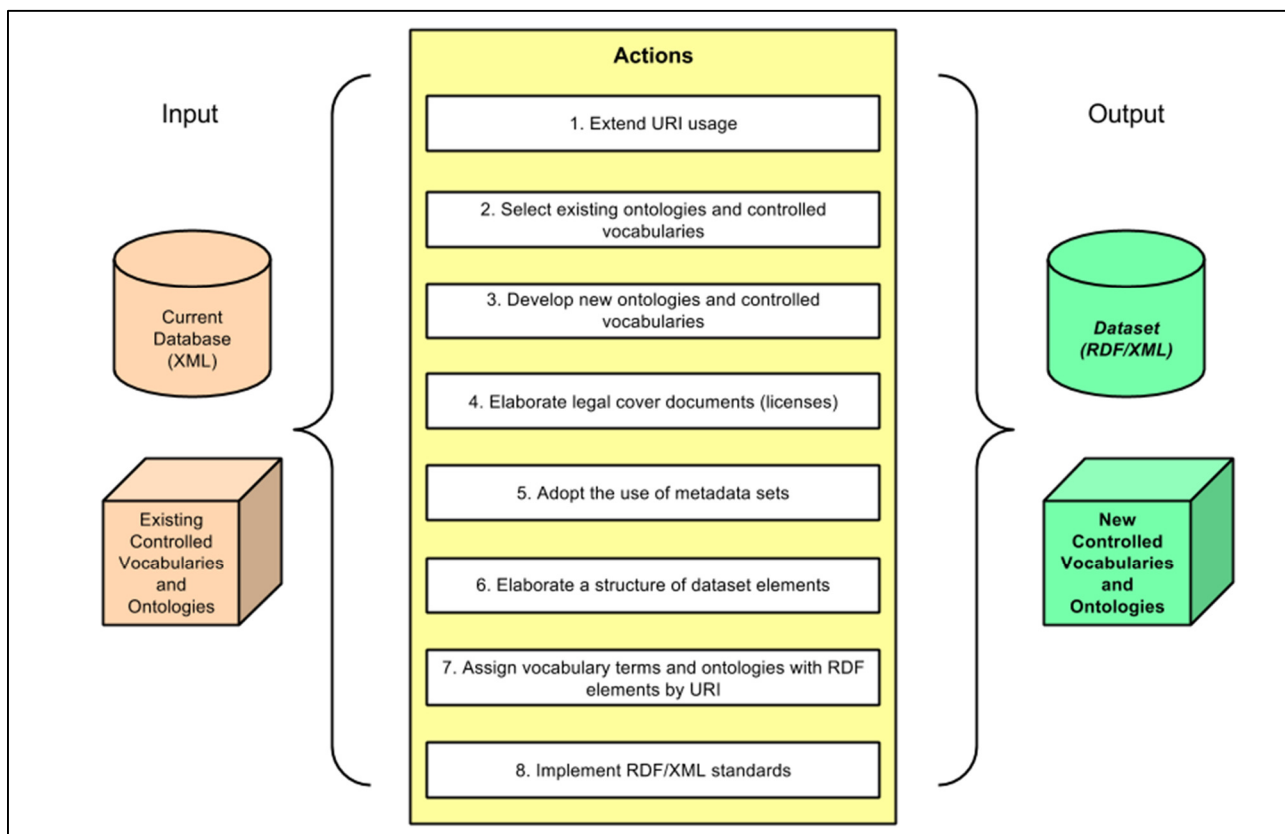
The elements 'Metadados' and 'Votacoes' are grouping elements with their value formed by a set of one or more elements; both with no attributes. The grouping element 'Metadados' has three elements; 'Votacoes' contains at least one or more elements 'Votacao'; and the element 'Votacao' has nineteen elements. None of the elements have attributes.

The grouping element 'Votos' (affiliated to the element 'Votacao') contains one or more elements 'VotoParlamentar', with no attributes. The element 'VotoParlamentar' has seven elements, also without attributes.

Results

From the analysis, eight actions were proposed to apply on existing data retrieval to publish government datasets in LOD (Figure 1).

Figure 1 - Synthesis of actions necessary for the development of government datasets



Source: the authors.

Previously to the implementation of these actions, it is important to be available information (input) about characteristics of existing databases and prior knowledge of available ontologies and vocabularies that may be part of the relationships and of the elements from new LOD dataset.

The actions identified in this study can be summarized in:

- Action 1: extend the use of URI identifiers for the identification of elements, attributes, and terms, explicit in additional documents, to facilitate dataset understanding and the rules of this system;
- Action 2: select ontologies and controlled vocabularies widely used by communities that can be useful to explicit relationships and elements of the new LOD dataset;
- Action 3: elaborate specific ontologies and vocabularies for relationships and elements that do not exist in ontologies and vocabularies adopted in Action 2;
- Action 4: develop additional documents containing legal coverage such as licenses of use and copyright, and link these documents with the dataset. It is important that these licenses be explicit in metadata (Action 5);

- Action 5: adopt the use of metadata elements sets of popular initiatives to inform more about datasets' content in the moment of data retrieval by external agents;
- Action 6: elaborate the logical structure of dataset, attributes, elements, values, and validation rules;
- Action 7: assign terms of selected ontologies and vocabularies to the dataset elements, linking them through URI, to extend the semantic load of these data;
- Action 8: implement RDF structures on XML markups, respecting the established forms for RDF/XML documents.

These actions should develop the LOD dataset (explicitly in RDF/XML format or equivalent) as output (result); and vocabularies and ontologies designed to meet the needs of the data context (Votações Nominais).

Conclusions

In the current form of dataset retrieval, it not considered important characteristics in the context of data, such as the use of controlled vocabularies and ontologies, directly interfering in independence between external agents and producers in data collection.

Although there are three metadata elements, there are no information about the content itself, such as author, license, source, date of creation, and date of publication.

The application of a recommendation model in this context served as a guideline for a development of actions proposed in this study, mainly by providing subsidies to elaborate the set of actions required for government datasets in LOD, allowing public managers to see important points that may be changed in available data that can be restructured into LOD datasets and, therefore, collectible by external agents.

It is expected that these actions applied on bills datasets will stimulate the application in other databases, in other spheres, and in other websites; but also stimulate an emergence of new Brazilian government datasets in this area.

References

The references was made following the ABNT rules.

BERNERS-LEE, T. The next Web. In: *TED2009*. Estados Unidos da América: fev. 2009.

BIZER, C.; HEATH, T.; BERNERS-LEE, T. Linked Data – The Story So Far: *International Journal on Semantic Web and Information Systems*, v. 5, n. 3, p. 1–22, 33 2009.

BOHMAN, J. *Public deliberation: Pluralism, complexity, and democracy*. [s.l.] MIT press, 2000.

BRASIL. *Lei No 12.527, de 18 de Novembro de 2011*. Regula o acesso a informações previsto no inciso XXXIII do art. 5º, no inciso II do § 3º do art. 37 e no § 2º do art. 216 da Constituição Federal; altera a Lei no 8.112, de 11 de dezembro de 1990; revoga a Lei no 11.111, de 5 de maio de 2005, e dispositivos da Lei no 8.159, de 8 de janeiro de 1991; e dá outras providências. Portal da Legislação, Brasília, 2011. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm>. Acesso em: 3 abr. 2015.

BRASIL. *Portal e-Cidadania*, 2015a. Disponível em: <<http://dadosabertos.senado.gov.br>>. Acesso em: 3 abr. 2015.

BRASIL. *Votações Nominais*, 2015b. Disponível em: <<http://dadosabertos.senado.gov.br/dataset/votaces-nominais>>. Acesso em: 3 abr. 2015.

BRICKLEY, D.; MILLER, L. *FOAF Vocabulary Specification 0.99*, 14 jan. 2014. Disponível em: <<http://xmlns.com/foaf/spec/20140114.html>>. Acesso em: 27 mar. 2015.

HEATH, T. *Frequently Asked Questions*. In: *Linked Data - Connect Distributed Data across the Web*, 2015. Disponível em: <<http://linkeddata.org/faq>>. Acesso em: 22 mar. 2015.

JENTZSCH, A.; CYGANIAK, R.; BIZER, C. *State of the LOD Cloud*, 2011. Disponível em: <<http://lod-cloud.net/state>>. Acesso em: 24 mar. 2015.

MALIN, A. M. B. Gestão da Informação Governamental: em direção a uma metodologia de avaliação. *DataGramaZero - Revista de Ciência da Informação*, v. 7, n. 5, out. 2006.

MARCONDES, C. H.; JARDIM, J. M. Políticas de Informação Governamental: a construção de Governo Eletrônico na Administração Federal do Brasil. *DataGramaZero - Revista de Ciência da Informação*, v. 4, n. 2, abr. 2003.

SANT'ANA, R. C. G.; RODRIGUES, F. A. Visualização de afinidades entre parlamentares mediante dados de votações no Senado Brasileiro. *Informação & Sociedade: estudos*, v. 23, n. 1, p. 49–59, jan. 2013.

SANTOS, P. L. V. A. DA C.; ALVES, R. C. V. Metadados e Web Semântica para estruturação da Web 2.0 e Web 3.0. *DataGramaZero - Revista de Ciência da Informação*, v. 10, n. 6, dez. 2009.

SANTOS, P. L. V. A. DA C.; SANT'ANA, R. C. G. Transferência da Informação: análise para valoração de unidades de conhecimento. *DataGramaZero - Revista de Ciência da Informação*, v. 3, n. 2, abr. 2002.

SCHMACHTENBERG, M.; BIZER, C.; PAULHEIM, H. Adoption of the Linked Data Best Practices in Different Topical Domains. In: MIKA, P. et al. (Eds.). *The Semantic Web - ISWC 2014*. Cham: Springer International Publishing, 2014a. p. 245–260.

SCHMACHTENBERG, M.; BIZER, C.; PAULHEIM, H. *State of the LOD Cloud 2014*. University of Mannheim, 2014b. Disponível em: <<http://linkeddatacatalog.dws.informatik.uni-mannheim.de/state>>. Acesso em: 25 mar. 2015.

SEMANTICWEB.ORG. SPARQL endpoint. In: *Semantic Web.org*, 2011. Disponível em: <http://semanticweb.org/index.php?title=SPARQL_endpoint&oldid=52541>. Acesso em: 04 jan. 2015.

W3C. *Data Licensing*. W3C, 2011. Disponível em: <<http://www.w3.org/wiki/index.php?title=TaskForces/CommunityProjects/LinkingOpenData/DataLicensing&oldid=49411>>. Acesso em: 20 mar. 2015.

W3C. *RDF 1.1 XML Syntax*. W3C, 2014. Disponível em: <<http://www.w3.org/TR/2014/REC-rdf-syntax-grammar-20140225>>. Acesso em: 22 mar. 2015.

W3C. *Linked Data Platform 1.0*. W3C, 2015. Disponível em: <<http://www.w3.org/TR/2015/REC-ldp-20150226>>. Acesso em: 24 mar. 2015.

Knowledge Organization and industrial heritage in São Paulo: the project Eletromemória



Vânia Mara Alves Lima
University of São Paulo
vamal@usp.br



**Marcia Cristina de Carvalho
Pazin Vitoriano**
São Paulo State University
marciapazin@marilia.unesp.br

Cristina Hilsdorf Barbanti
University of São Paulo
crishb@usp.br

Introduction

This paper discusses and analyzes the issues related to knowledge organization on industrial heritage, in the specific domain History of Electric Energy in the State of São Paulo, aiming to represent, retrieve and disseminate the information found in dispersed documentary collections, from its availability, in a web-based research tool.

These issues have arisen in the development of a thematic project, financed by FAPESP in two phases, aimed at collecting the historical heritage of the generation, transmission and distribution sector of electric energy in the state of São Paulo throughout the 20th century. In the first phase, named Eletromemória I, from 2007 to 2011, we carried out surveys and analyzes of the electric power generation, transmission and distribution units constructed during the decades of 1950-1970, representing the period of consolidation of the state electric system. In the second

phase of the project, named Eletromemória II, which began in 2012, about 50 units of electric power generation have been studied, among power plants and small hydroelectric plants (SHPs), built between 1890 and 1960. Interdisciplinary and inter-institutional, the project team, counts on professors, researchers and students from the three São Paulo universities - USP, UNESP and Unicamp - in the areas of History, Geography, Museology, Archival Science, Librarianship and Architecture.

Considering the various issues involved in the implementation of projects of this nature, Eletromemória II, by contemplating historical, documentary, environmental and industrial heritage aspects, defined four thematic axes of action, each of which addresses issues related to the area of knowledge referred to:

- **History:** focuses on the research on the transformations of productive units and their equipment, considering the history of technology, as well as the history of workers who experienced the operation of these plants at different times.
- **Industrial Heritage/Material Culture/Museology:** focuses on the research of the conditions of the assets represented by buildings, equipment and other elements of the material culture identified in the studied units, in a perspective of study of the remnants of the industrial heritage of the state of São Paulo and its museological potential.
- **Landscape/Environment:** based on a descriptive-comparative approach, we focus on the study of the landscape surrounding the places of implantation of the studied units, considering the original fauna and flora and their current conditions, the changes of the water regime and the history of use and occupation of the areas.
- **Information Science:** focuses on the identification of the documentary records related to the studied plants and the definition of parameters for the representation of the information related to the accumulated knowledge about each one of the analyzed units.

By aggregating different areas of knowledge, in an example of an interdisciplinary research community, the project requires a search instrument that consolidates the data obtained. In this work we approach the elaboration of the

Inventory of Industrial Heritage of Electric Energy in the State of São Paulo, an instrument that consolidates the research under two aspects: the terminological standardization and the standardization of the descriptive structure of contents.

As hypothesis, we consider that the mediation of knowledge organization systems in distinct communities can happen at the level of Documentary Languages and Terminology. In this sense, it is necessary to design tools that articulate and transpose knowledge domains and promote real communication bridges among its various agents.

The theoretical-methodological referential is found in the epistemological frontiers of Archival Science, Library Science and Museology gathered from the point of view of Information Science. This is an experience report, an exploratory qualitative approach.

Terminological standardization

The work of the Information Science axis group during Eletromemória I aimed at the elaboration of a controlled vocabulary to represent and retrieve the information stored in documentary collections belonging specifically to the domain of electric energy. The basic premise was that terminological control would guarantee the conceptual representation of the domain, even though the interpretation was submitted to the views of diverse communities and conditioned to the specific characteristics of the broad documentary typology of the focused universe. The controlled vocabulary was elaborated from the subject catalogs of archival, bibliographical and museological collections of Fundação Energia e Saneamento and had its first version included in the database Enerweb, available at: <http://www.energiaesaneamento.org.br/>.

Eletromemória II extends the mapping of documentary collections, and inserts new research axes, aiming to update the instrument of terminological control, including terms referring to the industrial heritage of the electric sector that will be used to describe the productive units of historical characteristic. We know that, especially since the 1990s, in the business environment of concessions, the regulation of the economic activity brought a minimum standardization of the terminological references of the collections. The existence of these minimum documentary standards to comply with the legislation allowed the interlocution among collections of different companies from the point of view of the used terminology. However, the occurrence of this type of

control for the collections of the beginning of the implantation of the electric power, between the end of 19th century and the middle of 20th century, is still extremely limited, lacking terminology control for the documentary production or for other technical assets issues, such as equipment and constructive techniques.

One of the characteristics of vocabulary control tools is also the possibility of terminological integration in different areas of knowledge. The association of research in both axes - information and documents - will enable the creation of a more comprehensive instrument, both from the point of view of the informed contents and the terminological control of the information, enabling the researcher to have ample and qualified access to the collection.

The field surveys carried out in the electricity generating units and in public and private collections aim to diagnose the archival, industrial, material culture and museological potential heritage, indicating their state of organization and conservation, as well as their relationship with the environment and its transformation in time, allow the collection of new terms to be inserted in the controlled vocabulary.

In order to include these new terms, a meeting was held with the researchers of the project for the elaboration of a conceptual map, as defined by Fendrich and Pereira (2006) and Rovira (2005), that is, propositions were organized about the main theme of the project (History of Electric Energy) which, visualized graphically allowed the understanding of the relationships among the connected concepts stimulating the solution of problems and enabling the perception of different paths within this conceptual system.

The discussions allowed to identify the terms: **Patrimônio industrial** (Industrial heritage); **Paisagem e Meio-ambiente** (Landscape and Environment); **Processos de eletrificação** (Electrification processes) and **Documentação** (Documentation) as subcategories of the category **História da Energia Elétrica** (Electric Energy History); to be restructured in the controlled vocabulary. All subcategories must be organized from the establishment of logical-semantic relationships among the terms that compose it. These relationships are established from the definitions of each term identified as belonging to that subcategory. The definition of a term should be drawn from what is prescribed by ISO 704 (2000) and 1087 (2000).

According to Dahlberg (1978), the definition of a term must state the essential characteristics of the concept, that is, the characteristics without which we cannot understand the concept that the term designates. The definition should also state the

common characteristics, i.e., those that allow two terms to be in the same subcategory/category, as well as the disjunctive characteristics, those that establish the boundaries of that subcategory/category.

The use of domain terminology as a reference for a controlled vocabulary ensures the control of the meaning of its terms and consequently guarantees the quality of the information it represents. In order to ensure future interoperability with other terminological control instruments, ISO 25964-1 (2011) is used for the preparation of thesauri and interoperability across vocabularies.

Thus, it is expected that the defined and adequately related terms, guarantee the meaning of the metadata that constitute the research tool Inventário do Patrimônio Industrial de Energia Elétrica de São Paulo (Inventory of the Industrial Heritage of Electric Power of São Paulo) and at the same time guarantee the content of the documents described there.

The Inventory of Industrial Heritage of Electric Energy in São Paulo

It is a research instrument that articulates the knowledge produced by distinct areas, referring to the same research object. After analyzing the concepts of each area, it was clear to the researchers that it was necessary to articulate the contents in order to develop a common data structure that met the minimum criteria of information representation of each one of the treated thematic axes, preserving its specificities. Therefore, the definition of the metadata structure should consider the existence of:

1. Informational needs for each area of knowledge;
2. Normative standards that meet these needs and that enable the articulation of contents on two levels: information of different nature on the same unit and information of the same nature on different units.

The performed comparative analysis took into account issues such as the great geographic and temporal scope of a project of this nature. Considering the universe of research, the researchers should combine information about units spread across all regions of the state of São Paulo, built at different times, and therefore with specific characteristics, whether constructive, functional or environmental. In addition, issues of administrative nature, such as the transition of ownership among various companies and government agencies, should also be considered in the construction of the instrument.

After analyzing the possibilities in the areas of Archival Science, Museology and Librarianship, it was realized that, for the purpose of the project, it would be interesting to use a descriptive instrument that could be applied to any of the units, describing it at several levels, which corresponds to the concept of multilevel description used by the international standard of archival description, ISAD-G (CONSELHO INTERNACIONAL DE ARQUIVOS, 2000).

Although it is a more comprehensive research tool than a traditional heritage guide, defined as a "documentary description tool that allows an overview of an archive service or a body and which presents practical information about the content and organization of each of the funds "(CUNHA, CAVALCANTI, 2008, p.183), or in librarianship perspective, "a document with instructions to guide users on the knowledge and exploitation of the collection of documentary bodies" (CUNHA; CAVALCANTI, 2008, p. 183), we realized that the concept could be applied to this case, and it is obviously necessary to broaden the concepts to be applied. The idea that each productive unit could be considered as a relatively independent unit, configured as an individualized set, made this option the most interesting one. In this way, it is possible to treat all the described collection from a documentary point of view, especially considering the research context in which the project is inserted.

At the same time, the final instrument would have to consolidate the concept of inventory used by the area of Industrial Heritage, in order to make it recognizable to all researchers. The record of goods and other things belonging to a person or community (RETTIG, 2009, p. 190) should follow the guidelines of Nizhny Tagil Charter, which indicates the need for field surveys, the elaboration of industrial typologies and site inventories that provide easy and free access to the public of interest. It also guides the inclusion of descriptions, drawings and photographs, as well as the indication of documentary sources (TICCH, 2003).

In order to meet the specificities previously presented, the instrument was divided into six large informational areas:

1. Fund/Collection - Identification of the production unit: containing basic identification and ownership data, construction data, geographic location, onset and end dates of operation. This area reconciles information from all the specific fields of knowledge, considering the accumulation and rights over the collection. The term Fund, used in

Archival Science, refers to issues similar to the term Collection, used in Library Science and Museology.

Within this item, we incorporate issues such as authorship, provenance and provenance of the collections, considering theoretical issues of Archival Science, but which were adapted to meet the methodological characteristics of the different areas.

2. History of the unit: with information on the history of the production unit, from the motivations for its creation, importance to the region and chronological data on the productive unit;

3. Industrial Heritage/Museology: with the description of the remaining physical structure (both architectural and equipment) and museological aspects, such as the analysis of the possibility of implementing a visitation route;

4. Landscape/Environment: with a comparative analysis of the environmental characteristics of the surroundings of the studied unit;

5. Archival collection: containing information on the location of documents of historical value, their conditions of preservation and access.

6. Keywords: all the research items underwent analysis and terminological standardization, constituting descriptors, formatted from the defined controlled vocabulary. In this way, terminological normalization is responsible for making content compatible among the different studied units, allowing the comparative analysis in different contexts.

Conclusions

Considering the hypothesis of this research, where terminology and information structure are responsible for the mediation of knowledge organization systems in distinct communities, the paper presented a report on the experience in which the development of the information representation instrument incorporates and articulates concepts from different areas of knowledge.

The comparative analysis of concepts used by the different areas on the History of Electric Energy in the State of São Paulo and the organization of the informational references on the theme allowed the creation of a descriptive instrument in which the terminological control and the registration of the contents are combined.

On the other hand, the elaboration of a single instrument, presenting information specific to each area of knowledge about the same object, articulated among them,

allows the researcher a comprehensive and interdisciplinary view of the researched unit, in this case, each small hydroelectric plant participating in the project .

We understand that, both the knowledge about each unit and the relations established among them, become clearer and can be used consistently by the various communities of researchers involved.

References

The references was made following the ABNT rules.

CONSELHO INTERNACIONAL DE ARQUIVOS. ISAD(G): *Norma geral internacional de descrição arquivística*: segunda edição, adotada pelo Comitê de Normas de Descrição, Estocolmo, Suécia, 19-22 de setembro de 1999, versão final aprovada pelo CIA. Rio de Janeiro: Arquivo Nacional, 2000.

CUNHA, M.B; CAVALCANTI, C.R.O. *Dicionário de Biblioteconomia e arquivologia*. Brasília, DF: Briquet de Lemos / Livros, 2008.

DAHLBERG, I. Teoria do conceito. *Ciência da informação*. Rio de Janeiro, v.7, n.2, p.101-107, 1978.

FENDRICH, L J; PEREIRA, L. Ensinar e Aprender no Ensino Superior Através de Mapas Conceituais In: SIMPEP, 8., 2006. *Anais*, Bauru.

INTERNATIONAL STANDARD ORGANIZATION. *ISO 1087: Terminology work - vocabulary*. Geneva : ISO, 2000

INTERNATIONAL STANDARD ORGANIZATION. *ISO 25964-1: Information and documentation - Thesauri and interoperability with other vocabularies – part 1 – Thesauri for information retrieval*. Geneva : ISO, 2011.

INTERNATIONAL STANDARD ORGANIZATION. *ISO 704: terminology work - principles and methods*. Geneva : ISO, 2000

RETTIG, J.M. *Glossário Sextilingüe de Patrimônio Industrial*. Santiago de Chile: Universidad Internacional SEK / TICCIH-Chile, 2009.

THE INTERNATIONAL COMMITTEE FOR THE CONSERVATION OF THE INDUSTRIAL HERITAGE (TICCIH). *Carta de Nizhny Tagil sobre o patrimônio industrial*. TICCIH, 2003. Disponível em: <http://ticcih.org/wp-content/uploads/2013/04/NTagilPortuguese.pdf> Acesso em 30/04/2015.

The Social, Cultural and Political Dimension of Knowledge Organization

Cultural Interoperability and Knowledge Organization Systems



Widad Mustafa El Hadi
University of Lille 3
widad.mustafa@univ-lille3.fr

1. Context and Rationale

This article addresses cultural interoperability in a context of rationalization. That means, it aims at the same time at taking a certain distance from rationalization questions and the challenges with which KO community is confronted. Examining cultural interoperability would involve other questions: how would it be possible to avoid falling into a downward technicization spiral since interoperability, especially the technical one, is rooted in the classic computing problems underpinning information systems compatibility from the point of view of their formats, data, interchange protocols and so on. What hinders semantic and consequently cultural interoperability is that the degree of success that can be achieved in the integration of multiple knowledge representation systems or knowledge organization schemes is constrained by limitations on the universality of human conceptual systems. For example, human languages do not all lexicalize the same set of concepts; nor do they structure (quasi-)equivalent concepts in the same relational patterns. As a consequence, even multilingual thesauri designed from the outset from the perspective of multiple languages may routinely include situations where corresponding terms are not truly equivalent (Hudon, 1997), (2005, 2012), (Green & al. 2002).

It is therefore essential to pay attention to the local, organizational, professional or national cultures, in KOS tools design and implementation, user studies, semantic Web technology and "intercultural" interfaces design. One of the questions we are addressing is, from where do culture of cultural interoperability grow? It will be interesting to examine intercultural exchange contexts: from technical interoperability to intercultural interoperability through semantic, military and enterprise interoperability.

2. Outline

The core issue in my presentation is Cultural interoperability in knowledge organization systems (KOS). Before examining this question I will first revisit the concept of interoperability and its layers with a special focus on semantic interoperability. This investigation is partly based on a previous paper (Favier & Mustafa El Hadi, 2013) where we have considered that semantic interoperability originated on one hand from KOS Translatability, and on the other hand from navigating within their diversity through interfaces linking one or two KOS.⁴ This will lead us to bring forth the following hypothesis: that semantic interoperability has a broad scope or coverage and that this very semantic interoperability is one of the core elements towards cultural interoperability. It should be noted however, that full semantic interoperability cannot operate in all contexts in spite of the efforts made in that direction. As a result there is even more to say about the difficulties ahead in trying to achieve “cultural interoperability”. The main question I will try to address here, is to what extent can this concept apply to KOS?

The article will be structured as follows: I will first review the ongoing research on KOS interoperability and define the concept of interoperability in its many layers (Miller 2002). I will then define “cultural interoperability” in its different aspects: within the military sphere, in diplomacy or in enterprise activities before trying to circumscribe it to KOS.

The term knowledge organization system (KOS) covers all types of schemes for organizing information and promoting knowledge management. They include classification schemes that organize materials at a general level (such as books on a shelf), subject headings providing more detailed access, and authority files which control different versions of key information (geographical names or personal names). This also includes less traditional schemes, such as semantic networks and ontologies (Hodge 2000). Because knowledge organization systems are mechanisms for organizing information, they are at the core of every library, museum, and archive. One

⁴ Our experience is based on the evaluation of the Visual Catalog (Papy, 2013) which links classification data using DDC, to RAMEAU authority file and other essential UNIMARC data. It is currently used by six French universities. The Visual Catalog is not based on translatability or semantic features but on an interface mapping between numerous vocabularies tailored for search and retrieval in education knowledge-domain. Two sub-domains vocabularies pertaining to schools and are enhancing mapping. We considered these two sub-domains as representing to two discourse communities or “cultures” the Visual Catalog is thus used as a hub linking the two vocabularies.

of the challenges we face in knowledge organization and representation is heterogeneity of systems both at the level of expressions and structure of conceptual content (see Green 2002, Doerr 2001; Chun & Moen 2001; Zeng & al. 2004, among many others). This point will be extensively developed in section 5.

3. The Importance of KOS Interoperability in Digital Environments

The pervasive power of digitization causes scientific, educational, economic and cultural communities to change their modes of accessing, sharing and disseminating knowledge. As a result this leads to a convergence between our cultural heritage, classic culture and technical culture. It is no surprise that some researchers have called for a “*digital humanism*” (Doueihy 2011), pointing out the way in which new technologies are becoming a sort of «culture» which “drives us into a new global cultural destiny” (Mustafa El Hadi, 2013). This concerns society as a whole. One of the driving forces of these developments is the scientific community i.e. academia which is both “maker” and “consumer” of these technologies.

General search engines on the Web are being so successful in performing the tasks they are designed for that they create an illusion of ease and efficiency that gradually shapes public opinion and influences decisions having to do with investment into information services, education, research or culture (DIGIKO, 2013).

Yet, the most important characteristics of the information environment speak of an increasing complexity of information discovery:

- information on the Web are shared globally (multilinguality and multiculturalism)
- Web is participatory in nature: experts and non-experts are involved in information production, information management and access
- information seeking scenarios are becoming multiple and unpredictable, with users who have different requirements
- the information sought is now heterogeneous, comprising text, images, sounds or datasets
- hybrid collections now contain both digital and non digital resources , with full text resources and metadata⁵ pointing to resources

⁵Metadata can be divided into four main categories not mutually exclusive and there is some degree of overlap between them according to its origins (Broughton 2010:50) professionally generated metadata, created by cataloguers and indexers using conventional controlled vocabularies

- author generated metadata, usually in the form of keywords added to the preliminaries of a paper

To date, most of the efforts in digital environment have been focused on technical interoperability issues: architecture, metadata standards, copyright and interoperability that are crucial to the management of digital repositories. Thus, the bulk of interoperability research has focused on technical and informational issues. Until recently, relatively little effort or expertise has been directed towards matters of semantic content access, semantic interoperability, or knowledge capture and facilitation. Even less attention is thus directed to KOS cultural interoperability apart from a few articles written by authors from KO community (Dahlberg, & Siwek, 1995; Miller 2002; Hudon, 2005; Fox & al. 1998; Favier & Mustafa El Hadi 2013). The issue is explored mainly in other discourse communities: in organizations (Bekkers 2005; Kousouris & al. 2011; Cayir & al., 2008) military and diplomatic domains (Clément, 2007, cited by Favier & Mustafa El Hadi (2013); Wimmer & al., 2008; Winslow & al. 2001). However there is a lack of general theoretical framework for considering these developments

KOS in general have long been used as a means of information exchange and has functioned as de-facto standard in knowledge organization and resource discovery: KOS are based on widely accepted and known scientific and educational consensus; as they are conceptual schemes they are language independent, widely used and available in many languages. Bibliographic classifications have some advantages over other controlled vocabularies such as thesauri, descriptor systems and subject heading lists, in that the notational encoding acts as a control device which is linguistically neutral. Retrieval problems such as the proliferation of synonyms, ambiguity of meaning, and other natural language factors can be avoided, making the exchange of information and cross-searching in a multi-lingual or multinational environment much easier. Classification process is at the core of all KOS whatever their structure and content is. For Brian Vickery “ Classification manifests itself, not only explicitly in the form of classification schemes, but also implicitly in other kinds of information languages which signal conceptual relationships between subjects, most notably thesauri and subject heading lists”, (Coates, 1988: 222). We believe that

-
- automatically generated metadata, derived from the resource itself using text mining or natural language processing tools, or harvested from the resource's existing metadata
 - user-generated metadata, attached by end users, usually on an informal and individual basis, but increasingly as a result of co-operative projects.

combining the two types of KOS (classifications & language-based KOS such as multilingual thesauri and vocabularies) could guarantee full interoperability.

Moreover, KOS are semantically rich terminology tools, traditionally used in library and bibliographic domains to enable information discovery - they are ubiquitous pathways to information in hundreds of thousands of document collections digital libraries, digital collections, and digital repositories. The full integration of KOS in digital environments would facilitate integration of the large corpora of recorded knowledge in heritage institutions and digitally-born resources on the Web. But most importantly, KOS can improve access to document contents in digital or hybrid collections and are indispensable in heterogeneous collections of documents (text, images, audio recordings, datasets, (Slavic, 2011; 2013)).

An overview of KOS has been presented by Hodge (2000), who grouped them into three general categories: (1) term lists, which emphasize lists of terms, often with definitions; (2) classifications and categories, which emphasize the creation of subject sets; and (3) relationship lists, which emphasize the connections between terms and concepts.

Many authors pointed out KOS heterogeneity and the challenges that their compatibility⁶ and interpretability represent. Heterogeneity characterizing Knowledge Organization Systems is embedded in their different structures and purposes even if they may vary in their structure, semantic, lexical, and notation or entry features (Iyer & Giguere, 1995, cited in Zeng & al. 2004). For example, they may cover different subject domains or have different scope and coverage; they may have semantic differences that are caused by variations in conceptual structuring (see also Soergel 2010)⁷; their levels of specificity and the use of terminology may vary; and the syntactic features, such as the word order of terms and the use of inverted headings, may be different. Such incompatibilities have presented problems in most mapping efforts. Initiatives for developing concordances or translation between a thesaurus and a classification or among various systems constructed on different principles are extremely challenging tasks. This is especially true when the target system has a

⁶ Early endeavors (1995) had been undertaken by ISKO to pave ways for making the KOS compatible, see Dahlberg, Siwek, (1995)

⁷ Soergel (2010: 12-36) suggested an approach to mapping between KOS including ontologies, classifications taxonomies thesauri and natural language based on deep semantics. One of the most interesting features of his work, because it is particularly relevant to our investigation, are the examples driven from Cross-language mapping and shades of meaning

higher level of specificity than the source system(s) (for a complete report see Zeng & Chan 2004: 386). But the most important distinction between the different kinds of KOS seems to be the diversity of semantic relations being displayed as noted by Green (2002: 15). In traditional classification systems, hierarchical relations and the relationship between synonyms and homonyms are the most important features⁸.

4. Defining Interoperability and its Layers

The term *interoperability* has many meanings, including the notions of communication, exchange, cooperation, and sharing of resources between systems. In fact, the essence of interoperability is that it is a *relationship* between *systems*, where each relationship is a manner of communication, exchange, cooperation and sharing (Carney & al. 2005).

This concept has an overall definition and more specific ones related to the different contexts in which it can be considered. Through the years, interoperability has tended to develop a broader scope and to relate to many contexts. I will now review the different definitions of the term and explore its scope. Paul Miller in his article “What is Interoperability? What Is It and Why Should I Want It?” (2002), noticed that the difficulty to define this concept: “Although not a perfect [definition] this comes closest to capturing my view of interoperability, and is informed by conversations over the past few years with people too numerous to name, all of whom have helped to refine the ideas behind these words: ***to be interoperable, one should actively be engaged in the ongoing process of ensuring that the systems, procedures and culture of an organization are managed in such a way as to maximize opportunities for exchange and re-use of information, whether internally or externally***”.

⁸ Green (2002, 15) pointed out the reasons of their heterogeneity and listed some of the solutions adopted to overcome this difficulty: “Their number exacerbates a problem that arises naturally from their individuality, that is, (some degree of) incompatibility. Responding to this situation, we have also developed a number of tools that show correspondences across classificatory systems, some of the same type, some of different types. Examples include: conversion tables between the Library of Congress Classification and Dewey Decimal Classification schemes correspondences within Library of Congress Subject Headings to equivalent classes within the Library of Congress Classification ; cross-language equivalences built into multilingual thesauri; the mapping of dozens of medical vocabularies to the Metathesaurus of the Unified Medical Language System (and mappings between versions of the WordNet lexical database But even where mappings across classificatory systems exist, we have no guarantee that the end result will lead to a unified semantic content and structure”

This general definition can be also compared to its scope in the enterprise context while referring at the same time to technical interoperability in its large sense: “Interoperability is the ability of disparate and diverse organizations to interact towards mutually beneficial and agreed common goals, involving the sharing of information and knowledge between the organizations via the business processes they support, by means of the exchange of data between their respective information and communication technology (ICT) systems”, in order to cooperate seamlessly, in an automated manner, in depth of time for a common objective” (Koussouris & al. 2011:1).

More definitions in the enterprise context where it is considered to be a major issue are suggested by the authors: “Enterprise Interoperability is perceived as a capacity of two or more enterprises, including all the systems within their boundaries and the external systems that they utilize or are affected by, in order to cooperate seamlessly, in an automated manner in depth of time for a common objective” (Koussouris & al. 2011: 1). The role of ICT in achieving interoperability in enterprise is also highlighted in Charalabidis & al. (2008, cited by Koussouris & al. 2011) the authors consider enterprise Interoperability as a “a field of activity with the aim to improve the manner in which enterprises, by means of Information and Communications Technologies (ICT), interoperate with other enterprises, organizations, or with other business units of the same enterprise, in order to conduct their business ». More interesting is their view of interoperability and its obvious links with semantic interoperability in KOS: “Consideration of the terms “Enterprise” and “Business Transaction” and decomposition of the enterprise concept in its major ingredients (e.g. people, assets, processes, knowledge, etc.); (.). Identification of the key EI challenges, as documented in the EI Research Roadmaps, (...); definition of a common EI taxonomy glossary in order to ensure common understanding of the key underlying terms. During this step, four different granularity levels for the EI taxonomy have been defined based on the prerequisites they require”

In the military and diplomatic fields, interoperability is defined in the US, NATO and Australia as “the ability of systems, units, or forces to provide services to and accept services from other systems, units, or forces and to use the services so exchanged to enable them to operate effectively together”. (Koussouris & al. 2001).

Donna Winslow put forward the reasons for considering cultural interoperability in military and diplomatic contexts. In a report submitted to the UN in 2000⁹ she gives the following definition: “ In order to promote civil military cooperation, it is important to understand some of the difficulties that can arise in peace operations between the members of these communities. (...) I will explore some of the tensions that can arise between the military and relief agencies. I have identified five possible points of tension to be found in peace operations, which I have been calling, a “cultural interoperability” model. These points of tension are related to organizational differences in terms of:

1. organizational structure and culture,
2. tasks and ways of accomplishing them,
3. definitions of success and time frames,
4. abilities to exert influence and control information,
5. control of resources. (see also with reference to a previous study”, (Winslow & al. 2001).

4.1. Interoperability in Knowledge Organization Systems

Closer to our concerns are the definitions and scope of the term suggested by Fox (Fox & al. 1998). The authors focused on interoperability in the context of Digital libraries thus identifying different layers likewise Miller (2002) who pointed out the different “flavors” of interoperability: “ Linking multiple systems in a useful way requires various levels of interoperability: Technical interoperability, which is concerned with hardware, networks, data types, and application compatibilities and protocols; Informational interoperability, which addresses content scope, language, metadata,

⁹ The report proposes that the UN should be able to deploy 5000 troops “as a brigade formation, not as a collection of battalions that are unfamiliar with one another’s doctrine, leadership and operational practices.” (UN 2000: 19) This is an important point. Outside of NATO militaries there is little commonality therefore UN operations run into problems due to a lack of compatibility or as I call it a lack of “cultural interoperability.” This can affect the UN’s ability to assure human security. In order to examine possible points of tension, which can arise in peace operations due to a lack of cultural compatibility between the various militaries in the UN operation in Bosnia, shad to develop a measure of "cultural Interoperability" . In an article wrote with her husband Colonel Peer Everts (Winslow and Everts 2001) she developed the concept of a cultural interoperability model. In UNPROFOR there was a great diversity in the organization of the operation and as my colleagues in COM would remind us, organizational studies show that diversity in organizations leads to increased communication and co-ordination problems and thus to potentially decreased organizational performance. We also demonstrated that one reason for the success of NATO in IFOR and SFOR was a common military culture shared between NATO participants in the operation. This commonality had been developed over fifty years of training and working together. Thus experience in careful planning, clear hierarchies and strong discipline ensured co-operation and co-ordination in diversified structures typical of multinational operations (Winslow & al. 2001).

naming conventions, semantics, and user Interfaces; Social interoperability which is concerned with personal and organizational rights and responsibilities" (...) It is clear that all levels of interoperability, including the social-political, must be addressed for a more tightly coupled global digital library to emerge".

Paul Miller (2002) gave a detailed account of the concept and identified different levels of interoperability calling them "Flavors of interoperability". He noticed the following: "Given such a wide scope within the suggested definition, it becomes useful to further subdivide the notion of interoperability". In his broad definition, Miller distinguished more layers in addition to the most common ones (semantic, syntactic and technical interoperability). He added political and human interoperability which has to do with the decision to make resources more widely available and has implications for the organizations: the inter-community interoperability; the legal interoperability and the international interoperability to which are related cultural issues. This involves usage practices, expectations, and requirements that vary from country to country. The layers added by Miller are quite relevant for anyone intending to integrate data from different locations.

We believe that semantic interoperability is the most important layer because it has an important impact on cultural interoperability: different types or levels of search and data interoperability, such as syntactic, functional and semantic, have been identified (Moen, 2001:2). He argues, "While syntactic and functional levels may focus on protocols used for information retrieval and communication systems, the semantic level concerns the understandings and meanings of interchanged data (...). In the hierarchy of challenges for effective interoperability, semantic interoperability is regarded as a grand-challenge research area for achieving high-quality interoperability". We will now see how semantic interoperability impacts cultural interoperability.

4.2. Semantic Interoperability

Semantic interoperability refers to the meaning of information to its human users, as opposed to the simple physical transfer of data. Interoperability at this level can fail if different users, or groups of users, use different terms for the similar concepts, or use similar terms to mean different things. Friesen (2002: 105) argues that "semantic interoperability is tied directly to communities of practice, and to the negotiation of meaning that occur within them." Because meanings assigned by people

and groups can vary over time and in subtle ways, semantic interoperability cannot be as easily ensured as technical interoperability but it can be improved. Attention must be paid to who the potential users of information systems are and what are the language that are necessary to communicate with them; there should be an agreement on the standard thesauruses and lists of terms that are to be used in metadata systems and there should be a consistent use of nationally existing coding systems ¹⁰

5. Review of Trends in Establishing Semantic Interoperability between Knowledge Organization Systems

The main foci of recent research activity into establishing semantic interoperability will be reviewed in this section with a special focus on multilingual thesauri.

The use of KOS to support more effective organization and retrieval of digital resources is not in itself a new concept. Web development in the mid-nineties was marked by numerous research projects building quality subject gateways and portals using classification and thesauri for browsing. The idea of using classification as a 'switching language' or 'pivot' to map indexing languages for the purpose of information integration and exchange was widely discussed as early as the 1970s (Marcella & Newton, 1995, cited in DIGIKO, 2013). This trend was especially strong in the UK where a number of SGs were developed within the Electronic Libraries Programme (eLib) (Kerr, 2009). As manual indexing has become increasingly difficult towards the end of the 1990s a number of projects explored automatic classification of web resources (GERHARD, SCORPION, CORC, Renardus). The evolution of the Web, from purely academic to commercial, business and social coincided with a significant improvement in technology. Web portal developers started to be primarily interested in the benefits of concept organization into practical and purposeful categories that their users/customers could easily combine or independently navigate. Subject orientated, simple hierarchical structures on gateways and portals have started to be combined with, or replaced by, a faceted organization of object/subject properties and attributes.

¹⁰ Guide for *Australian Flexible Learning Framework for the National Vocational Education and Training (VET) System*, produced by the Sharing Knowledge project, an initiative within the 2000-2004.

These applications led to the development of data formats and tools for the management of faceted vocabularies and their use in a Web interface (La Barre, 2006; Giess, Wild & McMahon, 2008, cited in DIGIKO, 2013). The wider exposure of tools such as thesauri and classification systems created a need for better understanding of the theory and principles underlying these tools. For instance, the lack of understanding of more sophisticated knowledge organization systems such as faceted classification resulted in web and other vocabulary standards being weak in supporting these systems (Slavic 2008; 2011, cited in DIGIKO, 2013).

Classifications are particularly well suited for use as a central mapping spine in connecting the vocabularies of different languages and seem to give visible results quickly and easily. This idea was explored by OCLC in terminology services (Vizine-Goetz et al. 2004). With the emergence of MARC formats for classification data, ISO standards for structured vocabularies and especially since the creation of SKOS which initiated most of copyright protected vocabularies to be released as linked data on the Open Web we can see a start of a new stage of KOS use in the digital environment. Over past five years many of these vocabularies were published as linked data on the Web and the main obstacle for their wider and quicker deployment is lack of expertise. It makes economic sense that the knowledge and experience associated with these tools are made more universally accessible and usable by both peoples and programs across sectors and domains. This is the first step towards the increasing participation and interaction of different knowledge communities in improving KOS terminology and enhance integration and linking of knowledge landscape. Moreover, publishing of KOS on the Web opens many possibilities of exploitation of KOS in many fields of activities in which they were not so easily available. Knowledge organization as a field and KOS as tools can thus be able to operate within the whole universe of recorded knowledge thus providing a unique semantic layer for knowledge discovery. Dextre Clarke & Zeng (2012: 23) showed how is SKOS vital in such endeavors: "The W3C Recommendations SKOS is designed to support publications of vocabularies such as thesauri on the Web. And at its heart is a data model that explicitly distinguishes between concepts and labels to represent concepts. (...). In addition to all that is conveyed by SKOS Core for relationships between concepts, the extension provides additional support for identifying, describing, and linking lexical entities", Dextre Clarke & Zeng (2012: 23), (see also Dextre Clarke, 2011)

To this end, over the past ten years a number of W3C workgroups in the area of Semantic Web development have showing a keen interest in knowledge organization systems traditionally used in the bibliographic domain (thesauri and classifications) and their power in capturing and communicating meaning. This is especially the case with W3C development groups in the area of vocabularies, web ontology standards, linked data and more recently the W3C Library Linked Data Incubator Group. This sparked research activities, projects and forums such as Networked Knowledge Organization Systems (NKOS) that promote collaboration between bibliographic and other domains in the area of subject retrieval.

The wider exposure of tools such as thesauri and classification systems created a need for better understanding of the theory and principles underlying these tools. For instance, the lack of understanding of more sophisticated knowledge organization systems such as faceted classification resulted in web and other vocabulary standards being weak in supporting these systems (Slavic 2008). Vanda Broughton (2010: 50) highlighted the proximity of SKOS with the faceted classifications and their relevance for the semantic Web: “In order to interact fully with the Semantic Web, a faceted classification must be made visible (Slavic, 2008, cited by Broughton, 2010), and all of the elements that make faceted KOS particularly useful for online organization and retrieval must be visible too. Some existing Semantic Web applications, such as Simple Knowledge Organization System (SKOS) exhibit features similar to the faceted KOS, and indicate that faceted KOS may provide a model for the elements in Semantic Web formats, as well as the terminologies to inhabit them”.

In Mustafa El Hadi (2013), I investigated the ways in which universal classifications, as language independent concept schemes, can assist humans and computers in structuring and presenting information and formulating queries. A wider application of universal knowledge classification in digital humanities may contribute not only to better resource discovery but could also help in improving terminological and research tools for knowledge transfer between humanities and other fields of knowledge.

5.1. Mapping between Multilingual Vocabularies

Establishing of equivalence is essential when creating multilingual subject vocabularies or merging multiple vocabularies. According to the assumption that all languages are equal in a concordance, there exists a question of whether views of a

particular culture that are expressed through a controlled vocabulary or a classification can be appropriately transferred during the mapping (Zeng & al. 2004)

The ideal match is the establishing the “one-to-one” relationships between terms in different vocabularies and different languages, but this often proves elusive (Zeng & Chan 2004). To map vocabularies or create multilingual or multi-disciplinary vocabularies other difficulties crop up when trying to establish equivalence between these KOS. Different linguistic expressions for the same concept, different degrees of specificity and polysemic terms are hindering these attempts.

The ISO 5964 standard specifies the following types of equivalence relationships: exact equivalence; inexact equivalence; partial equivalence; single-to-multiple term equivalence; non-equivalence (ISO, 1985). Similar attempts at defining term equivalence reported for by (Zeng & Chan 2004) include the degrees of equivalence suggested by Riesthuis (2003) who has defined them as follows: exact equivalence, single-to-multiple equivalence, non-equivalence, loan terms, and inexact equivalence.

Although there are complex challenges to achieve semantic interoperability, many efforts have been made into this direction in order to support high quality information exchange and integration across communities (Chen, 1999; Chung & Moens 2001; Zeng & Clarke, Hoolande & al. 2013; Doerr, 2001. Chung & Moens (2004) identified two primary areas in semantic interoperability research: data attributes and data values. The data attribute area addresses questions as to the names, labels, semantics, and granularity of metadata scheme elements and database fields. The data value area addresses the data or information provided in an element or database field. Their research is based on Doerr (2001) approached the semantic interoperability issue with a concept-based mapping between numerous vocabularies for subject retrieval illustrating the relationship between two dimensions—data attribute and data value in the context of semantic interoperability. He noted that as the semantic interoperability of the two dimensions increases, the overall semantic interoperability increases proportionally. To improve the capability for information exchange or integration, and to improve the search capacity across these resources, it is essential to enhance interoperability of systems and data. This involves addressing not only interoperability at the syntactic and functional levels but at the semantic level (Moen, 2001). Various information services such as federated searching, harvesting metadata, and gathering are often dependent on the capability and quality of semantic

interoperability across different collections, systems, domains, and communities (Arms et al., 2002; Chun & Moens 2004).

Chung & Moens (2004) highlighted the need for semantic interoperability to meet users' expectations of discovering relevant networked information resources not only via searches across heterogeneous as well as within homogeneous systems and collections (Chen, 1999; Tennant, 2001, cited by Chung & Moens (2004)). This is the reason why semantic interoperability is considered as in the "hierarchy of challenges for effective interoperability", Chun & Moens (2004:2). We believe therefore that the most important step towards effective interoperability is conditioned by the semantic layer

To make different KOS compatible, many efforts have focused on term or concept mapping. Both intellectual and machine-based efforts to map among vocabularies in the online era have been reported for by Zeng and Chan who categorized them in the following groups (Zeng, 1992; Zeng & Chan 2004). The authors reported on many initiatives aiming at unifying KOS in an integrated source or interface. Efforts at establishing interoperability among KOS in different languages and in different structures can lead to the integration, mapping, and creation of knowledge organization systems for information sharing in a networked environment as Zeng & Chan (2004) indicated. These projects vary in both the targets for mapping and in the methods used (for an extensive description of methods, KOS categories, (see Zeng & Chan 2004). The authors described the projects in terms of the languages and structures involved. The efforts in establishing interoperability among existing KOS in different languages are examined and grouped in two categories: the first, consisting of MACS and Merimee, attempts to establish links among the vocabularies or authority files involved. The second, consisting of the remaining examples, utilizes a switching mechanism among the vocabularies involved. Ten projects had have been reported for (Zeng & Chan (2004: 379-380). As for establishing interoperability among different existing KOS, the projects described have been categorized by the authors according to the structural types that the involved KOS possess. For an extensive description of the projects see Zeng & Chan (2004: 380-381). Regarding the methodological analysis, they analyzed both the conventional and the new ones used in mapping such as derivation/modeling; translation/adaptation; direct mapping; co-occurrence mapping; switching; linking through a temporary union list; linking through a thesaurus server protocol. They also analyzed methods used in link storage and management.

They comprise: authority records, concordances, semantic network and lexical database (2004:386).

In their conclusion the authors confirmed the following: interoperability of KOS is an unavoidable issue in today's networked environment. In addition to efforts to achieve interoperability among existing KOS for specific collection(s) and project(s), collection-independent KOS, especially vocabulary knowledge bases, are highly desirable. The need for reconciling different subject vocabularies in networked environments is indisputable. Results from recent efforts to achieve interoperability among different vocabularies are encouraging. These projects have identified and experimented with a variety of methods. The question remains as stated by them (2004-386): "Have we fully exploited technological capabilities in our efforts to improve subject access to the myriad resources now available in the networked environment?" Their analysis and findings are extremely valuable for our investigation.

In a recent paper, Hoolande (Hoolande & al. 2013: 464) explored the feasibility of using subject vocabularies as a linking hub to the Semantic Web. They noted that the semantic enrichment and integration of heterogeneous collections can be facilitated by using subject vocabulary for cross linking between collections, since major classifications and thesauri (*e.g.*, Library of Congress Subject Headings [LCSH], Arts and Architecture Thesaurus [AAT], DDC, Répertoire d'Autorité-Matière Encyclopédique et Alphabétique Unifié [RAMEAU]) have been made available following Linked Data principles. Reusing these established terms for indexing cultural heritage resources represents in their view a big potential for LAM (Library, Museums and Archives). For this purpose they examined the feasibility of using subject vocabularies as a linking hub to the Semantic Web in advance of such efforts. They adopted a semi-automated vocabulary alignment technique used earlier by Doerr (2001), Isaac et al. (2008)¹¹.

So here is, a non exhaustive summary, compiled from the various methodologies used in the literature, that are used to achieve semantic interoperability

¹¹ Doerr (2001) and Isaac & al. (2008) identified four general approaches toward vocabulary reconciliation or alignment: (1) lexical alignment techniques, (2) structural alignment, (3) extensional alignment, and (4) alignment using background knowledge (Hoolande & al. 2013: 465-466)

- 1) mapping between numerous vocabularies (Doerr, 2001)
- 2) concept-based mapping between numerous vocabularies for subject retrieval Doerr (2001)
- 3) classifications used as a central mapping spine in connecting the vocabularies of different languages being used (Marcella & Newton, 1995; Vizine & al. 2004)
- 4) using subject vocabulary for cross linking between collections using subject vocabularies as a linking hub to the Semantic Web (Hoolande & al. 2013)
- 5) establishing a variety of links among the vocabularies or authority files involved (Zeng & Chan, 2004)
- 6) using a switching mechanism among the vocabularies (Zeng & Chan, 2004)

5.2. Multilingual Mapping: the Case of Multilingual Thesaurus

Why do I focus on thesauri? The reason is that this tool represents a category of KOS built on natural language. They are increasingly multi-lingual and capable of accommodating complex cultural demands. The major challenge of multi-lingual thesauri, making controlled vocabulary work across language boundaries (Smiraglia 2012). Examining this KOS category is one of the key elements for considering cultural interoperability. This idea will be developed in section 7.

Heterogeneity between different thesauri is attributed to the following, as identified in Doerr (1996; 2001) and Dachelet 1997: different word uses, due to different natural languages; different coverage, due to different states of development, different scope and varying user needs; different semantics, due to different conceptualizations. This will occur typically between thesauri from different languages, but it may also be due to different aspects of classification; different semantic relations are often due to the enforcement of mono-hierarchies, but they can also be due to different classification aspects (Doerr, 2001).

Doer (2001) observed that any kind of relations between terms from thesauri in different natural languages are referred to as translations. In his opinion, translation *per se* differs from concept-based mapping and cross-concordances in significant ways. To this end he preferred referring to the typology suggested by Roland Dachelet (1997, cited by Doerr, 2001). This author proposed distinctions between different kinds

of multilingual thesauri, defining “translated thesauri” as thesauri where each concept is optimally interpreted in words of another or multiple languages, to allow speakers of those languages to understand better and use the concepts of this thesaurus more effectively. He noted that such translations are in general not established indexing terms of the target language; he then defined correlated thesauri as an aggregate of multiple thesauri consisting of established indexing terms (concepts) of the respective user communities and a set of concept-based mappings between the concepts from the different thesauri of that aggregate; then the Interlingua Thesaurus was defined as a thesaurus made of concepts created by fusing each cluster of similar concepts from different social groups into a new concept. The Interlingua provides a sharing of concepts between social groups. It may not contain any of the original concepts of any user group but it contains a set of compromises to remove interpretational differences. Its concepts may again be translated and correlated to other thesauri.

Zeng & Chan (2004) reviewed a set of methodological options for multilingual mapping. In order to meet users’ requirements in networked environments, a major decision facing those attempting to establish interoperability among knowledge organization systems is the choice of an appropriate method. The first question is whether to integrate, map, or create a system. The possible options are similar to those suggested by Riesthuis (2003, quoted in Zeng & Chan (2004)) with regard to different approaches in creating multilingual thesauri: translation, merging, and creating from scratch.

Dagobert Soergel¹² summarized the main limitations of existing KOS taking as an example the AGROVOC thesaurus. This can be seen as complementary to Hudon (1997, 2005; and Doerr 2001; Zeng & al. 2004) He mentioned successively:

-Lack of conceptual abstraction: thesauri and other traditional KOSs are collections of terms (generic or domain-specific), ordered in a poly-hierarchic lattice structure or a mono-hierarchic tree structure and interlinked with some very broad and basic relationships.

-Limited semantic coverage: most thesauri do not differentiate concepts into types (such as *living organism*, *substance*, or *process*) and have a very limited set of

¹² Dagobert Soergel, Boris Lauser, Anita Liang, Frehiwot Fisseha, Johannes Keizer and Stephen Katz “Reengineering Thesauri for New Applications: the AGROVOC example (available at <http://www.fao.org/docrep/008/af234e/af234e09.htm#TopOfPage>)

relationships between concepts, distinguishing only between hierarchical relationships, i.e. NT/BT, and associative relationships, i.e. RT.

-The concept relations provided by most thesauri force all relations into two broad categories, hierarchical and associative.

-Lack of consistency: since the relationships in thesauri lack precise semantics, they are applied inconsistently, both creating ambiguity in the interpretation of the relationships and resulting in an overall internal semantic structure that is irregular and unpredictable..

-Limited automated processing: traditionally thesauri were designed for indexing and query formulation by human beings and not for automated processing. The ambiguous semantics that characterizes many thesauri makes them unsuitable for automated processing.

In a survey on multilingual thesauri, Michèle Hudon (1997: 89-90) noticed the following problems:

The stretching of a language to make it fit a foreign conceptual structure to the point where it becomes barely recognizable to its own speakers; The transferring of a whole conceptual structure from one culture to another, no matter whether it is appropriate; and The literal translation of terms from the source language into meaningless expressions in the target language.

According to Hudon (1997-90): "true equality for all languages in a multilingual thesaurus has a better chance of being achieved if the following global requirements are met: the thesaurus should be built within a semi-centralized administrative structure, with representatives of each language/culture on the decision-making team; all linguistic versions of the thesaurus should be developed simultaneously from the ground up; the thesaurus designers should be native speakers of the language in which they work, with a good knowledge of the other languages involved; distinct term banks should be built independently for each language with terms found in source language documents; identity and symmetry of structures are not required across the various linguistic versions of the thesaurus, and single-to-multiple equivalence, "orphans", and variations in hierarchies, etc. are allowed; the use of neologisms should be very restricted if allowed at all; thesaurus development software which allows for nonidentity of descriptor records and for rotation of source and target languages should be used; physically separate displays for each language represented should be produced".

7. Cultural Interoperability

The concept of Culture is disciplinary and describes in general the various phenomena that make up the collective beliefs and activities of a certain group of people. Discussions of culture refer generally to shared values, history, language, collective memory, social attitudes, preferences and practices (Beghtol 2002a:509).

In its early definitions cultural interoperability is related to the military and diplomatic fields which are considered as organizational bodies in the way we mentioned above. The term can be traced back¹³ in Winslow and Everts writings (2001). "In an article I wrote with my husband Colonel Peer Everts (2001:19) I developed something called a "cultural interoperability model" in order to examine possible points of tension, which can arise in peace operations due to a lack of cultural compatibility between the various militaries in the UN operation in Bosnia. In UNPROFOR there was a great diversity in the organization of the operation and as my colleagues in COM would remind us, organizational studies show that diversity in organizations leads to increased communication and co-ordination problems and thus to potentially decreased organizational performance."

Sotirios Koussouris, Fenareti Lampathaki, Spiros Mouzakis, Yannis Charalabidis, and John Psarraas suggested a broader definition of cultural interoperability. Their definition applies to the enterprise and business contexts and should perhaps be seen as strongly limited to these contexts: "Interoperability between organizations, people and enterprises systems that have different languages and different cultural aspects such as Politics, Religion, regional Art, Traditions and Social Customs defines the concept of *Cultural Interoperability*. As such, Cultural Interoperability is the degree to which knowledge and information is anchored to a unified model of meaning across cultures. Enterprise systems that take into consideration Cultural Interoperability aspects can be used by transnational groups in different languages and cultures with the same domain of interest in a cost-effective and efficient manner. Cultural interoperability mechanisms are based on the assumption that both high-level and low-level layers of Enterprise reflect culture and that the linguistic encoding of knowledge and information is therefore culturally biased. These interoperability mechanisms and considerations address the ability of

¹³ According to King (2006), Bernard Boene was the first author who coined this term

enterprises to understand and co-manage context from any source and of any kind, therefore realizing the cooperation between enterprises with major cultural differences”

7.1. Cultural interoperability in Knowledge Organization (KO)

As I mentioned in the introduction a few articles have dealt with cultural interoperability in the Knowledge Organization community. In a recent article (Favier & Mustafa El Hadi, 2013) we examined two possible dimensions of semantic interoperability: one based on KOS Translatability, the other one, on navigating within their diversity, exploiting links between heterogeneous KOS for information discovery and retrieval purposes¹⁴. In other words we tested in the Visual Ctalalog (Papy, 2013) what is called the approach of a “one-stop” seamless search process (see also Zeng & Chan 2004), when entering one query the user can retrieve results from various heterogeneous KOS. This assumption lead us to put forward the following hypothesis: semantic interoperability is the most fundamental step towards cultural interoperability.

7.2 The concept of Translatability

Our hypothesis is built on translation possibilities between different languages, called translatability in linguistic communities. I will examine here the scope of the concept of “translatability ” of and how it applies to KOS. From the translation community, Raquel de Pedro (1999: p. 546-559) accounted for two points of view when considering translatability: the universalist one and the monadist one. Supporters of the former approach claim that the existence of linguistic universals ensure translatability. Those who endorse the latter approach maintain that each linguistic community interprets reality in its own particular way and the universalist approach jeopardizes translatability . The approach, that has become known as the Sapir-Whorf hypothesis is generally not applied in its strongest form, since this would imply the impossibility of effective communication between the members of different linguistic communities (see also Green, 2002).

¹⁴ This is a major issue when considering convergence of information sources, the Web, digital libraries, archives, museums and repositories and the need for multidisciplinary and multicultural information. Web scale discovery tools improving search and retrieval such as OCLC or EBSCO federated search shows the directions to follow today. This technology allows the simultaneous search of multiple searchable resources. A user makes a single query request which is distributed to the search engines participating in the federation. The federated search then aggregates the results that are received from the search engines for presentation to the user.

From Knowledge Organization community, Rebecca Green showed how translatability is linked to semantic and lexical universals. “The extent to which we can satisfactorily translate between human languages is, it would seem, a good measure of the degree to which natural language incorporates semantic and lexical universals. (...) In enumerating conceptual universals that must be accommodated by any comprehensive knowledge organization or knowledge representation system, we may productively look to the semantic and lexical universals found in natural languages” (Green 2002: 20).

Green, Bean, & Hudon (2002: 311) examined whether a concept’s hierarchical level affects the likelihood of its universality across schemes for knowledge representation and knowledge organization. The authors drew empirical data on equivalents from a bilingual thesaurus, a pair of biomedical vocabularies, and two ontologies. Their findings show that “conceptual equivalence across resources occurs significantly more often at the basic level than at subordinate or superordinate levels. Attempts to integrate knowledge representation or knowledge organization tools should concentrate on establishing equivalences at the basic level. The implications of semantic and lexical universals survey showed the existence of conceptual universals in knowledge organization and knowledge representation systems” (for a complete account of this survey see (Green 2002; Green et al. 2002). It appears to the authors that there are dozens of conceptual units (e.g., sun, bad, think) that are more or less universally apprehended. There are also some numbers of conceptual abstractions (e.g., predicate-argument structure, countability, animacy, hierarchy) that are used universally in structuring our conceptual systems (...) where specific conceptual units are shared, we should construct cross-system mappings.

What would justify characterizing some conceptual content as a necessary feature of all human language? Green referred to Immler (1991) who took up a variety of meanings purported to exist in all human (he chose 18 languages) — for example, the distinctions between light and dark and between alive and dead— and then suggests “no human society could afford not to be able to express these ideas in its language” (pp. 39-40), Cited by Green (2002: 18). “cognitive contents can be universal across all cultures and all languages— they will be universal to the extent that the real world around us , which is the same for humans all over the world , is the object of cognition” argues Immler.

This idea resonates with the assumption according to which, in the communication process, as well as in the case of interoperability, perception and cognition play a leading role in building, structuring, and disseminating human information. Human communication begins when someone wants to transmit information in their mind to someone else. As such, cognitive models are basic elements of human-to-human communication (Brodeur et al. 2003)

Green, Bean, & Hudon, (2002: 311-312), in their seminal paper suggested that “conceptual units shared by many languages are more likely to occur at the basic level than at hierarchical levels above or below the basic level” suggest that the “best we can do in integrating two knowledge-based schemes is to construct such crosswalks as are truly appropriate and then rely on the individual structures of the schemes being “integrated.” Knowledge of these conceptual universals can at least facilitate the very considerable work that knowledge integration will always require”.

The assumption that meanings are given in a language, and that subjects can be constructed by a universal set of semantic universals given by natural languages has been questioned by Hjørland (2003: 109) when he posited that: “The basic assumptions and attitudes have been that the basic units of KO are semantic relations between concepts. Such semantic relations cannot primarily be established by universalistic assumptions, but much primarily be understood as domain specific, as uncovered by (and constructed by) scientific disciplines”. But we believe however, that our assumption makes sense while attempting to build general framework for describing what cultural interoperability in KOS is. Linking the two points of view (translation theory ¹⁵ and techniques) and (KO theoretical findings) represents another avenue available for research in semantic and cultural interoperability.

7.3. Fundamentals of cultural interoperability for KOS

An analysis of the literature demonstrates that in order to build a theoretical framework to account for cultural interoperability in KOS some basic concepts are to

¹⁵ In his own experience as a translator made von Humboldt perceive the need for translation: “translation is one of the most necessary tasks of any literature” (in Schulte & Biguenet, 1992: 56, quoted by Pedro). According to his hypothesis, the structural differences which exist between languages are no obstacle for translation. The reason that von Humboldt proposes to explain this is that each linguistic community has a potential of expression which can generate resources for verbalizing every extra-linguistic area, even those which go beyond its own social and cultural experience

be conveyed/ brought up. I will suggest first considering the concept of “universals in KO” as a basis for considering both semantic and cultural interoperability

7.4. From semantic Interoperability to Cultural Interoperability

Beghtol (2002a : 507) wrote: “new technologies have made the increased globalization of information resources and services possible. In this situation, it is ethically and intellectually beneficial to protect cultural and information diversity”. She analyzed in her study the problems of creating ethically and globally accessible and culturally acceptable knowledge representation and organization systems and foundation principles for the ethical treatment of different cultures established on the basis of the United Nations Universal Declaration of Human Rights. From an ethical point of view she mentioned the need for a global and local access to information in any language, available at any location at any time and for any purpose for any individual, culture, ethnic groups or domain. She suggested that the concept of “cultural hospitality” can act as a theoretical framework for the ethical warrant of KOS.

The two concepts, “Cultural Warrant”, and “Cultural Hospitality” are relevant for exploring cultural interoperability had been defined in (Beghtol (2002b: 45). Cultural warrant postulates that: “a knowledge organization system is more likely to be useful and appropriate for those who are members of a culture and that it is less likely to be useful and appropriate for those who belong to some different culture, at whatever level of society that culture or domain may reside. Thus, a knowledge organization system that is appropriate for the elements of one culture may not recognize elements that are highly important for some other culture, and such exclusions pose problems because we need to integrate knowledge across cultural, geographic, and linguistic boundaries. Important tensions thus exist between culturally specific knowledge organization systems and the need to provide systems that are both globally accessible and culturally appropriate. In general, knowledge organization systems for global usage would need to incorporate all the various syntactic and semantic foundations of any and all of the world’s cultures, and the creators of knowledge organization systems need to create techniques for poly-cultural information retrieval. Since the foundations of different cultures may be in conflict with each other, we need to develop theories and techniques for incorporating any possible cultural assumptions that might be used for information retrieval”

The concept of “Cultural hospitality” complements and extends the principles of cultural warrant. Both concepts are essential for a theoretical framework for KOS that will privilege the needs and existence of different cultures, whether they are national, ethnic, domain or disciplinary cultures: “The concept of “Cultural hospitality” has long been established as one of the desiderata for the notation of bibliographic classification systems. Hospitality is the ability of a notation to admit new concepts appropriately and to accommodate them in the correct relationships with other concepts. To develop a new theoretical framework for knowledge organization systems, it is helpful to expand the concept of hospitality by 1) including provisions for hospitality beyond notational issues and 2) broadening hospitality to include different cultures as well as new concepts. These extensions of hospitality provide us with a new concept, “cultural hospitality”. This concept implies that knowledge organization systems should be “permeable” to different points of view and different cultural attitudes and practices”, (Olson 1996, p. 9 quoted by Beghtol 2002b: 47). Although these two concepts are relevant, however their implementation may not be easy as Beghtol observed. She suggests the need for discussing, assessing and testing its potential for effective implementation.

The considerations sketched above can open a path of research that until now has been largely focused on technical and semantic interoperability. What hinders semantic interoperability and consequently cultural interoperability is that the degree of success achievable in the integration of multiple knowledge representation systems or knowledge organization schemes constrained by limitations on the universality of human conceptual systems. These limitations have been observed in the case of multilingual thesauri (see above section 5.2.). Natural human languages do not all lexicalize the same set of concepts; nor do they structure (quasi-)equivalent concepts in the same relational patterns (....) As a consequence, even multilingual thesauri designed from the outset from the perspective of multiple languages may routinely include situations where corresponding terms are not truly equivalent (Hudon, 1997, 2005), Green et al. 2002)

It is obvious that the lack of equivalence in mappings across knowledge organizations schemes severely impacts retrieval scenarios. The limitations are due to the nature of concepts at the basic level, where concrete entities are more likely to accept mapping through different languages as suggested in Green (2002: 23) “The basic level concept arose in the context of classifying physical objects. It is not always

clear how to apply some of the converging processes noted above, especially those related to perception and function, to abstract concepts, processes, events, and so forth (...) We will assume, however, that the basic level concept does apply to these less concrete contexts and that the several linguistic criteria that tend to converge for concrete entities will likewise converge for non-concrete entities. The confirmation of our hypothesis that full equivalence between tools occurs more often at the basic level than at either superordinate or subordinate levels for both knowledge representation schemes and knowledge organization schemes should illuminate future efforts to build universal tools of subject or semantic access. Specifically, it suggests that crosswalks between such tools should focus on mappings at the basic level, without attempting to impose a comprehensive mapping at all hierarchical levels”

7.5. Domain Analysis as an Element of a General Theoretical Framework for Examining Cultural Interoperability

Domain Analysis in its broad scope can possibly be adapted to a general framework addressing cultural interpretability in KOS. In his paper on “Epistemology of Domain Analysis, in *Cultural Frames of Knowledge*”, Smiraglia (2012: 111-112) noted that “Domain Analysis is at the heart of KO, for without it we would have no ontological matter to constitute our KOS. Culture and activities” (...) The domain is best understood as a unit of analysis for the construction that is a domain is a group with an ontological base that reveals an underlying teleology, a set of common hypotheses, epistemological consensus on methodological approaches, and social semantics”. The concept of “domain” and the scope of Domain analysis I am considering here is based on Hjørland and Albrechtsen (1995, p. 400) hypotheses, extended by Smiraglia (2012); and Tennis (2003).

Hjørland and Albrechtsen (1995, p. 400) define domains as “thought or discourse communities, which are parts of society’s division of labor” (..). Smiraglia (2012:111) noted that “A domain can be seen, according to these authors, as a type of discourse community”. Domain Analysis definition and scope had been extended and developed by Smiraglia (2012) and Tennis (2003: 191). Tennis considers the definition of “domain” as prior to defining DA definition. In reference to Hjørland and Albrechtsen text, he wrote the following: “(..), the term domain is not the same as discourse community. [They] go on to review the literature that uses the concept of domain under many terms. They cite “speciality/ discipline/domain/environment”

(Hjørland and Albrechtsen, 1995 p. 401) as the unit of study. As a consequence, the definition and its boundaries are muddied. It can be noted that for Hjørland (1995, 1998, 2002), it seems more important to define the *domain analytic paradigm* than the object of inquiry, the domain. And it is this problem of definitional muddiness I address here. There are a number of concepts similar to domains in Information Science; like Communities of Practice, and Epistemic Communities. It seems apparent to the casual reader what a domain might be. It could be an area of expertise, a body of literature, or even a system of people and practices working with a common language". Tennis added what he called "the two axes of DA" essential in his view to consider when analyzing a domain: "There are two axes of consideration when analyzing a domain: Areas of Modulation and Degrees of Specialization. Areas of Modulation set the extension of the domain and Degrees of Specialization set the intension. Each of these axes has two parameters. Areas of Modulation must state 1) the totality of what is covered in the domain analysis – the extension and 2) what it is called – its name. The Degrees of Specialization must 1) qualify the domain – state its focus and 2) state where the domain is positioned against other domains – its intersection (2003: 194)

If we sum up the basic elements to consider when building a KOS we will realize that the ingredients for a successful semantic and cultural interoperability between two or more categories of KOS (a classification scheme, vocabulary, thesaurus, ontology, dealing with the same knowledge-domain are embedded in the very definition of "Domain" (see Tennis, 2003 (191-192) for an extended definition of Domain, Hjørland 1993, 1995; 2002). A domain has: a boundary, a specific terminology, basic units concepts, terms, semantic relations, classification schemes and a shared ontology.

Hjørland (2002: 422) developed the importance of the role of domain studies in Information Science. While this discipline originates from special librarianship and documentation it lacked from general librarianship. He argues: "the core idea of the specialist library approach may be that information resources should be identified, organized and communicated to serve specific goal". (...). In my opinion even general librarianship has to cope with different domains and may well benefit from considering the domain analytical view. One cannot treat all domains as if they were fundamentally similar, and a theoretical approach to Library and Information Science should consider different discourse communities".

Thesaurus is a domain specific tool built for describing, indexing and retrieval. These specific KOS are built on all natural languages, in different cultures and

different organizations, sometimes within the same culture or language. Their importance in multilingual contexts is not any more to be proved. Doer (2001:1) wrote: “Thesauri are created in different languages, with different scope and points of view and at different levels of abstraction and detail, to accommodate access to a specific group of collections. In any wider search accessing distributed collections, the user would like to start with familiar terminology and let the system find out the correspondences to other terminologies in order to retrieve equivalent results from all addressed collections. Hjørland showed in his article (2002:425) the relationship between specific domain vocabularies exemplified by the thesaurus and his theory of Domain Analysis: “thesauri are mainly domain specific vocabularies, and the methodology of designing them can be seen as one (implicit) form of domain analysis.”

In Mustafa El Hadi (2008: 303) I mentioned the importance of reference tools, such as terminologies and thesauri for specialized information and communication. I wrote: “One of the purposes of information and communication sciences is the construction and sharing of meaning. This process involves building reference tools such as thesauri for indexing and retrieving an increasing digital data. Meaning construction through these tools is based on well known conceptual methods. To guarantee communication between the various actors of a field of knowledge, or what we can call a ‘community of experts’ which uses its own discourse these tools should be rigorously and coherently designed”. I defined in the same paper (2008: 302) “discourse community” as “ a concept describing all forms of communication that contribute to a particular, institutionalized way of thinking; and ‘community’, which in this case refers to the people who use, and therefore help to create a particular discourse. (...) If one wants to become a member of a certain discourse community, it requires more than learning the lingo. It requires understanding concepts and expectations set up within that community”. This point of view had been described by Smiragilia (2012: 116) as a domain analytical as pertaining to Domain Analytical approach in knowledge organization.

8. Conclusion

Semantic interoperability has a broad scope and we consider it as one of the core elements towards cultural interoperability. It should be noted however, that full semantic interoperability cannot operate in all contexts in spite of the efforts made in that direction and that there is even more to say about the difficulties ahead in trying

to achieve “cultural interoperability”. What hinders semantic interoperability and consequently cultural interoperability is that the degree of success achievable, in the integration of multiple knowledge representation systems or knowledge organization schemes, is constrained by limitations on the universality of human conceptual systems. We believe that combining the two types of KOS (classifications & language-based KOS such as multilingual thesauri and vocabularies) could progress towards full interoperability. Web ontology standards, such as SKOS (Simple Knowledge Organization System) is designed specifically with the intention to provide a common data model for sharing and linking knowledge organization systems via the Semantic Web. Standards such as SKOS are also meant to be used as a vehicle for publishing and sharing knowledge organization systems that were not born digital. The considerations sketched above can open a path of research that until now has been largely focused on technical and semantic interoperability. However there is a lack of general theoretical framework for considering these developments. We suggest that cultural interoperability and its framework should become a new epistemological perspective in KO.

Acknowledgements: The author is deeply grateful for the interesting discussions engaged during the DIGIKO (DIGITAL Culture and KOS) project writing. Special thanks to Sylvie Davis, Laurence Favier, Andrea Scharnhorst, Philip Mayer, Fabrice Papy Aida Slavic, and Peter Stockinger of the DIGIKO project consortium. Many ideas in this paper also originate from our discussions in writing the call for papers for the Special issue on “Cultural Interoperability”, *Communication*, (University of Laval) with the co-editors, Laurence Favier and Dominique Vinck (forthcoming). The author thanks also Gérard Prunier, a political analyst and an independent international consultant on geopolitical affairs, for the fruitful discussions engaged about cultural interoperability in diplomatic, political and military fields.

References

- Arms, W. Y., Hillmann, D., Lagoze, C., Kraft, D., Marisa, R., Saylor, J., et al. (2002). A spectrum of interoperability. *D-Lib Magazine*, 8(1), <http://www.dlib.org/dlib/january02/arms/01arms.html>.
- Brodeur, J, Edwards, G., Bédard, Y., Bernard Moulin, B. (2003), Revisiting the Concept of Geospatial Data Interoperability within the Scope of Human Communication Processes, In *Transactions in GIS*, Blackwell Publishing, 2003, 7(2): 243–265

Chung, E., Moens, W. (2007), "The Semantics of Semantic Interoperability: A Two-Dimensional Approach for Investigating Issues of Semantic Interoperability in Digital Libraries, Proceedings of the American Society for Information Science and Technology, http://www.unt.edu/wmoen/publications/Chung_Moen_ASIST_Short_Final.pdf

Beghtol, B. (2002a), "A proposed ethical warrant for global knowledge representation and organization systems", *Journal of Documentation*, Vol. 58 Iss 5 pp. 507 – 532

Beghtol, B. (2002b), "Universal Concepts, Cultural Warrant and Cultural Hospitality" in *Advances in KO*, Proceedings of the Seventh International ISKO Conference 10-13 July 2002 Granada, Spain Organized by the ISKO - Spanish Chapter and The University of Granada, p.45-49

Broughton, V. (2010), Emergent Controlled Vocabulary, in *LCN*, (Mustafa El Hadi & Hudon (ed), Vol. 6-N°3/2010 :50)

Carney, D.R., Hall, J.A., LeBeau, L. (2005). Beliefs about the nonverbal expression of social power. *Journal of Nonverbal Behavior*, 29, 105–123.

Coates, E.J. (1988), "The role of classifications in information retrieval : Action and thought in the contribution of Brian Vickery, *Journal of Documentation*, Vol. 44 Iss 3 pp. 216 – 225
Permanent link to this document: <http://dx.doi.org/10.1108/eb026826> Downloaded on: 03 August 2015

Dagobert Soergel, Boris Lauser, Anita Liang, Frehiwot Fisseha, Johannes Keizer and Stephen Katz Reengineering Thesauri for New Applications: the AGROVOC Example available at <http://www.fao.org/docrep/008/af234e/af234e09.htm#TopOfPage>

Dextre Clarke, S. (2011) "ISO 25964: a standard in support of KOS interoperability", in A. Gilchrist and J. Vernau (eds), *Facets of Knowledge Organization, 4-5 July 2011, London*. London, Emerald. <http://www.iskouk.org/conf2011/papers/dextreclarke.pdf>.

Dextre Clarke, S., Zeng, M.L. (2012), "From ISO 2788 to ISO 25964: The Evolution of Thesaurus Standards towards Interoperability and Data Modeling", in *Information Standards Quality*, Special Edition: Year in Review and State of the Art of the Standards, p.20-26.

Doueïhi, M. (2011). *Pour un humanisme numérique*. Paris: Seuil.

Doerr, M. (2001). Semantic problems of thesaurus mapping. *Journal of Digital, Information*, 1(8). <http://jodi.tamu.edu/Articles/v01/i08/Doerr/>.

DIGIKO (2013) Pré-Projet DIGIKO, un projet de recherche (Digital Cultures and Universality in KO) proposé avec quatre partenaires (la France en tant que pilote, le Royaume-Uni, les Pays Bas et l'Allemagne, conjointement avec un panel d'experts internationaux. Une version actualisée a été présentée durant la conférence *Knowscape*, à La Hague, le 23 octobre 2013.

Favier, L. and Mustafa El Hadi, W. (2013) "L'interopérabilité des systèmes d'organisation des connaissances : une nouvelle conception de l'universalité du savoir ?", in F. Papy (ed), *Problématiques émergentes dans les sciences de l'information*. Paris, Hermès.

Fox, E.A. Marchionini, G. (1998), " Toward a Worldwide Digital Library, in *Communication of the* ACM April 1998/Vol. 41, No. 4, p.29-32

Friesen, N 2002. Semantic Interoperability and Communities of Practice. Global Summit Conference, Adelaide. <http://www.educationau.edu.au/globalsummit/papers/nfriesen.htm>

Green, R. (2002), Conceptual Universals in Knowledge Organization and Representation, in *Advances in Knowledge Organization*, in *Advances in KO*, Proceedings of the Seventh International ISKO Conference 10-13 July 2002 Granada, Spain Organized by the ISKO - Spanish Chapter and The University of Granada, p.15-27.

Green, R., Bean, C.A., Hudon, M. (2002) Universality and Basic Level Concepts (2002), in *Advances in KO*, Proceedings of the Seventh International ISKO Conference 10-13 July 2002 Granada, Spain Organized by the ISKO - Spanish Chapter and The University of Granada, p.311-317.

Hjørland, B. (2003) Fundamentals of Knowledge Organization, *Knowl. Org.* 30(2003)No.2

Hjørland, B. (2008) What is Knowledge Organization (KO)? *Knowl. Org.* 35(2008)No.2/No.3

Hjørland, B. (2002), "Domain analysis in information science: Eleven approaches –traditional as well as innovative", *Journal of Documentation*, Vol. 58 Iss 4 p422-462

Hodge G., (2000), "Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files", *The Digital Library Federation Council on Library and Information Resources* <http://www.clir.org/pubs/abstract/pub91abst.html>

Hoolande, S.V. De Wilde, Verborgh, Mannens, De Walle, & Hercher, (2013) "Evaluating the Success of Vocabulary Reconciliation for Cultural Heritage Collections", *JASIST*, 64(3):464–479, 2013

Hudon, M. (1997). Multilingual thesaurus construction: Integrating the views of different cultures in one gateway to knowledge concepts. *Knowledge Organization*, 24, 84–91.

Hudon, M. (2005). Accès thématique en bibliothèque numérique : le rôle du langage documentaire de type « thésaurus ». *Les Bibliothèques numériques*, sous la dir. de Fabrice Papy, 151-177. Paris : Hermès Science.

Hudon, M. (2012), ISO 25964 : pour le développement, la gestion et l'interopérabilité des langages documentaires, in : *Documentation et bibliothèques*, n° thématique Normalisation et normes pour la gestion de l'information et des documents, vol. 58/3, 2012, 130-140.

Immler, M. (1991). "Is semantics universal, or isn't it? On the relation of language, thought and semantic structure". In D. Zaefferer (Ed.), *Semantic Universals and Universal Semantics*, 37-59. Berlin, New York

Dahlberg, I. Siwek, K; (editors) (1995), *Compatibility and Integration of Order Systems*, Research Seminar Proceedings of the TIP/ISKO Meeting, Warsaw, 13-15 September 1995, organized by The Polish Library, the Society for Professional Information Association.

Miller, P 2002, Interoperability. What is it and Why should I want it? *Ariadne* Issue 24. <http://www.ariadne.ac.uk/issue24/interoperability>

Koussouris, S. Lampathaki, F., Mouzakitis, S., Charalabidis, Y., Psarraa, J., (2011), "Digging Into The Real-Life Enterprise Interoperability Areas Definition And Overview Of The Main Research Areas" available at <https://www.academia.edu/2852773/>

Mustafa El Hadi (2008), "Discourse Community Analysis Sense Construction Versus Non-Sense Construction", in *Advances in Knowledge Organization*, Vol. 11 (2008) "Culture and Identity in Knowledge Organization", (Tennis & Arsenault editors), Proceedings of the Tenth International ISKO Conference, 5-8 August 2008 Montréal, Canada, p.302-306.

Mustafa El Hadi, W. (2013), "Faceted Classifications as a Model for Universal Classifications". In: From Libraries to the Web, *SRELS Journal of Information Management*, Vol. 50, (6); December 2013; p721-734.:

Papy, F. (2013), "Classification and visualization: augmenting user independence and enhancing collections use", in Proceedings of UDCC Seminar, *Classification & Interfaces to Knowledge*, The Hague, 23-24 October 2013.

Pedro, R.D. (1999) "The Translatability of Texts: A Historical Overview", *Meta : journal des traducteurs / Meta: Translators' Journal*, vol. 44, n° 4, 1999, p. 546-559.

Slavic, A. (2011) "Classification revisited: a web of knowledge", in A. Foster and P. Rafferty (eds), *Innovations in information retrieval: perspectives for theory and practice*. London, Facet. pp. 23-48.

Slavic, A. (2008). Faceted Classification: management and use. *Axiomathes*, vol. 18, p. 257-271.

Slavic, A. (2011). Classification revisited: a web of knowledge. In: *Innovations in information retrieval: perspectives for theory and practice* . Ed. by A. Foster and P. Rafferty P. London: Facet. pp. 23-48.

Slavic, A (2013) Classification in knowledge transfer between languages, domains and collections in humanities: from legacy collections to the Web of linked data, Presentation at European Seminar, Ecole Doctorale, 12th April, 2013, Lille

Smiragilia, R. (2012) "Epistemology of Domain Analysis". in *Cultural Frames of Knowledge*", Smiragilia & Lee editors), Ergon Verlag, pp 111-124

Soergel, D. (2010) "Concepts in context: proceedings of the Cologne Conference on Interoperability and Semantics in Knowledge Organization, July 19th-20th, 2010, Ergon Verlag Publisher, 13-36

Tennant, R. (2001). Different paths to interoperability. *Library Journal*, 126(3), 118-119.

Tennis, J. T (2003): Two Axes of Domains for Domain Analysis, *Knowl. Org.* 30(2003) No.3/No.4 191-195.

Vizine-Goetz, D., Hickey, C., Houghton, A., Roger, Th. (2004) "Vocabulary Mapping for Terminology Services", *Journal of Digital Information*, 4 (4).

Wimmer, A. (2008), "Elementary strategies of ethnic boundary making", in: *Ethnic and Racial Studies*, 31:6, 1025 —1055.

Winslow, D. and Everts, P. (2001) "It's not a Question of Muscle: Cultural Interoperability for NATO". In NATO - The First Fifty Years. G. Schmidt, ed., Hampshire, UK: Palgrave/MacMillan pp. 85-104.

Zeng, M.L., Chan, L.M., (2004) "Trends and Issues in Establishing Interoperability Among Knowledge Organization Systems, *JASIST*, 55(5):377–395, 2004.

Reflections on the development of a methodology for subject analysis



Roberta Cristina Dal'Evedove Tartarotti
São Paulo State University
roberta_tartarotti@yahoo.com.br

Paula Regina Dal'Evedove
Federal University of Pernambuco
paula.dalevedove@ufpe.br



Mariângela Spotti Lopes Fujita
São Paulo State University
fujita@marilia.unesp.br

1 Introduction

Investigations around the indexer as a professional reader provide important observations about the human mind as well as considerations for the improvement of subject analysis procedures. In view of this, one of the great challenges for Knowledge Organization and Representation as a scientific and social discipline of reflection and theoretical production (GARCÍA MARCO, 1995), confluent to the understanding of the problems that involve registered knowledge, focuses on the creation of methodologies aimed at the indexing process that promote the intellectual processing of subject analysis.

The theoretical and methodological scenario of Indexing, as an area of study and research, presents guidelines that enable the procedures of analysis, synthesis

and representation of the informational content of documents in the professional practice of the indexing librarian. Despite the advances, some issues remain open in the theoretical and normative field of this process of thematic information organization, as is the case of subjectivity in understanding concepts, which has been present in the discussions of the area for a long time and accompanies the indexer throughout his/her professional trajectory. The notoriety of subjectivity in the indexing process stems from its concatenation with several other social, cultural, political, physical, linguistic, logical and cognitive aspects, applicable to the understanding of subjects and conditioning the result of indexation.

Implicated for its theoretical and applicable dimensions, subject analysis, in the indexing practice, stems from comprehension processes permeated by subjective acts derived from prior knowledge and from professional training and experience, as well as from the cultural and ideological context of the indexing librarian and the information system. Thus, the development of methodologies aimed at subject analysis of effective control of the elements that permeate the understanding of concepts is a challenge to be faced by Knowledge Organization and Representation. Fujita (2013, p. 156) argues that the nature of the indexing process, although "[...] empirical and subject to intuitive applicability, demands scientific investigations aimed at improving the professional knowledge of those who execute it, as well as the instruments, methods and techniques used to do so ". However, from which theoretical and methodological aspects is it possible to circumscribe the subjective actions of indexing librarians who work in information contexts permeated by sociocultural aspects?

Due to the relevance covered by the theme, we work with the repertoire of subject analysis regarding the understanding of concepts, seeking to make considerations about the socio-cultural references present in professional actions. The intention in this study is not to provide answers to the raised problem, but to reflect on the challenges surrounding the creation of tools for subject analysis, given the cultural diversity with which the indexing librarian deals with in the indexing practice of contemporary information domains .

2 Understanding concepts in the indexing practice

Documentary reading is part of the activities performed by the indexing librarian, considered the initial phase that triggers all the other stages of the indexing process and which involves three main aspects: understanding the content of the document,

identification of the concepts that represent this content and selection of valid concepts for retrieval in information systems - overlapped during the reading of the document (FUJITA, 2003b, p. 64; TARTAROTTI, 2014, p. 147).

Giasson (1993) advocates the interactionist view of reading as a process of communication and interaction between three interrelated variables: text-reader-context. In subject analysis, the greater the interaction of these variables during documentary reading, the greater the level of professional understanding about the subject of the document. In general, the variable *text* represents the textual structure in documentary reading and contains the ideas of the author, in which textual knowledge is part of the reader's previous knowledge and provides conditions of understanding when he/she is able and familiar with the different typologies and textual structures, which facilitates understanding the subject. However, "the reception of the information obtained from the text through the reading takes place in a different way from reader to reader, through the way each one assigns meaning to its content" (DIAS, NAVES, 2007, p. 47)

The interaction between the indexing librarian and the text as a professional reader occurs through the use of metacognitive strategies and his/her background knowledge, keeping his/her objective of representing the text for future retrieval in mind and also considering the limitations of the activity and the objectives of the informational domain in which it is inserted (FUJITA, 2003a). This interaction consists of an amplitude of coordinated procedures and perceptual, linguistic and cognitive operations, in which the characteristics of the textual structure provide support to the indexing librarian's knowledge, increasing the understanding of the approached subject.

Understanding does not depend only on the intrinsic characteristics of the text, but on the background knowledge shared between the author and the reader, in which the author possesses, in his/her conception, a representation of the world and to understand a text is to relate elements of that representation. Reading is a practice that aims to give perceptual and intellectual meaning to the text. In this context, reading includes the reader and his/her subjective capacity to interpret. Thus, the reader is the second variable of the interaction process with objectives for documentary reading, being the element of greater influence and able to interfere in documentary representation.

In turn, the variable *context* refers to the subject analysis in specialized information contexts and all influential elements at the time of professional reading, i.e., "[...] all the conditions in which the reader is when he/she enters in contact with a text "(GIASSON, 1993, p. 40). In Van Dijk's (1997) conception, it is the set of properties of the social situation, which are relevant for the production, understanding and/or functioning of discourse and its structures. Given this, the context represents not only the professional background knowledge and the objectives of the activity in question, but also the environment in which this professional activity is carried out in documentary reading.

To make the meaning of the document's information content less ambiguous as possible, the document must be analyzed according to the context in which it is inserted, as this variable tends to add value to certain concepts, which assume new meanings only by changing the analysis environment (PINTO MOLINA, 1993). Therefore, the concept only makes sense when considering the situation context, in which the indexing librarian must adjust his/her decisions according to the need required by the medium.

In fact, the stages of identification and selection of concepts should be directed to the process of understanding and should be investigated from the perspective of other scientific fields, in which the study of documentary reading becomes interdisciplinary when considered as a practice of giving perceptive and intellectual sense to the text by the actions of the indexing librarian and his/her subjective capacity to interpret, as well as the specificities of the textual typologies and the informational domain in which the subject analysis occurs.

Background knowledge/professional cognition is part of the socio-cognitive context as a mental representation of the reader. Hjørland's (2002) theoretical view of the proposal of interaction between the subject and the social environment is twofold: the indexing practice is an intellectual process that depends on cognition and; the context domain contributes to the identification and selection of concepts that cover the users' needs and the enunciation environment.

3 Challenges for the construction of a methodology for subject analysis

There are important theoretical and applicable initiatives that provoke reflections and throw light on the implementation of a methodology for subject analysis that

surrounds the main aspects that permeate the indexing practice. However, research on subject analysis tends to check indexer processing patterns (DIAS, NAVES, 2007), that is, it seeks to identify strategies capable of extracting a subject from its document - focus on the process.

Despite advances in the specialized literature, very little is known about the actions of the indexer and his/her influence on the final result of the indexing process, especially when knowledge organization problems in the contemporary world are admitted, many of them linked to the "[.. .] difficulties in treating documents in the face of the manifestation of social, cultural and political diversities" (LARA, 2013, p. 238). Therefore, research on the subject analysis phase that covers questions about human action from objective social reality is still incipient. Dal'Evedove and Fujita (2013, p.44) understand that this "paradigm shift is necessary for understanding the professional reality in specific information domains", considering that improvements and better specifications in the development of indexing practice are emerging needs that deserve a closer look in the area.

In this sense, further research is important in the area of Knowledge Organization and Representation that seeks to measure the true impact of the indexer's particular actions on the indexing practice, linking the characteristics of sociocognitive approach with the sociocultural aspects that surround the informational context. This is due to the understanding that subject analysis cannot be studied separately from the social context in which indexing is performed. In addition to the questions related to the understanding of concepts in indexing practice and which are directly related to the indexer as a professional reader, research aimed at the theme needs to discuss about new investigative possibilities that allow to recognize the informational needs of the users in the face of different information domains.

In this regard, Carvalho (2004, p. 165) argues that "[...] it is contradictory to offer services and products to meet the informational needs of users that are not really known, nor in their minimalist characteristics." However, despite demanding efforts, recognizing the user as a guiding element in the indexing practice is the best alternative to overcome deficiencies and improve the quality of the generated informational products, aiming at increasing use. Complementing the issue, the author points out that the provision of ready-made products and services, especially the ones "[...] tested only from the perspective of the librarian, as something immutable or susceptible only of minor adaptations" must be abandoned in library practices.

It is observed that subject analysis is conditioned to certain objectives, organizational contexts and to the information retrieval system and, therefore, in the indexing practice socio-cognitive and sociocultural issues intertwine. In this sense, in order to address some theoretical and normative gaps that permeate the subject indexing, it is considered relevant to refer researches through the socio-cultural approach, specifically when the objective is to create methodologies for subject analysis. This is because, "the value of information is dependent on the meaning attributed by the subject, which implies conceiving information in its enunciation context to link it to the specific culture of each knowledge domain" (DAL'EVEDOVE, 2014, p. 116).

In this way, addressing sociocultural issues in subject analysis is a salutary path for the area, due to the need to reflect on the indexing librarian's performance in relation to a professional activity carried out in various informational domains. It is therefore accepted that the sociocultural bias applied to the studies of subject analysis is an important alternative to confront the contemporary configurations of the area of Knowledge Organization and Representation, through the possibility of acknowledging the social modes of thought of indexing librarians and users, social actors who present different opinions, impressions, behaviors and personal and collective interests.

4 Conclusion

In view of the presented discussion, accepting and recognizing that the indexing activity is based on socio-cognitive and sociocultural issues, it is a necessary path for research on the theme. To investigate indexing practice under the bias of the sociocultural approach is to consider that, in addition to the problematic about the process, issues related to social actors and informational domain are effectively included, aiming to contribute with theoretical and methodological subsidies that seek to fill some of the remaining deficiencies in Indexing.

The observation of cognitive, cultural and social aspects that permeate the indexing practice, relativized, is a way of considering the relationships that the social actors establish with the informative contents present in the documents. In this sense, it is understood that, for the development of a methodology for the subject analysis, the referenced aspects need to be considered.

In addition, the improvement of indexing practice is directly related to the development of methodologies for subject analysis, because from this the professional identifies and selects concepts representative of the document, which will be translated into descriptor terms through an indexing language. Therefore, research on the theme that is routed under the sociocultural bias will be in line with that advocated by Saracevic (1999), for which the studies undertaken in Information Science must integrate the various manifestations and behaviors of the information phenomenon (information analysis) and, in parallel, recognize the behaviors, effects and interfaces within a context of social dimensions.

Thus, the present research encompasses key issues present in the scope of Indexing of theoretical-conceptual and practical-applied dimensions, intertwined by continuous changes in the specific information domains in the midst of the composition and recomposition of society in the contemporary world, discussions still dispersed, with understandings to be consolidated.

References

The references was made following the ABNT rules.

CARVALHO, I. C. L. *A socialização do conhecimento no espaço das bibliotecas universitárias*. Rio de Janeiro: Interciência, 2004. 185 p.

DAL' EVEDOVE, P. R. *O tratamento temático da informação em abordagem sociocultural: diretrizes para definição de política de indexação em bibliotecas universitárias*. 2014. 266f. Tese (Doutorado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2014.

DAL' EVEDOVE, P. R; FUJITA, M. S. L. Estudo sociocultural da comunidade discursiva do tratamento temático da informação em bibliotecas universitárias. *Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação*, v. 18, n. 36, p. 23-50, jan./abr., 2013.

DIAS, E. W.; NAVES, M. M. L. *Análise de Assunto: teoria e prática*. Brasília: Thesaurus, 2007. 116 p.

FUJITA, M. S. L. A importância teórica e prática da indexação na fundamentação científica da organização e representação do conhecimento. In: DODEBEI, V.; GUIMARÃES, J. A. C. (Org.). *Complexidade e organização do conhecimento: desafios de nosso século*. Rio de Janeiro: ISKO-Brasil; Marília: FUNDEPE, 2013. p. 147-159.

FUJITA, M. S. L. *A leitura documentária do indexador: aspectos cognitivos e linguísticos influentes na formação do leitor profissional*. 2003. 321 f. Tese (Livre-Docência em Análise Documentária e Linguagens Documentárias Alfabéticas) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2003a.

FUJITA, M. S. L. A identificação de conceitos no processo de análise de assunto para a indexação. *Revista Digital de Biblioteconomia e Ciência da Informação*, Campinas, v. 1, n.1, p. 60-90, 2003b.

GARCIA MARCO, F. J. Los contenidos y la secuencia de la organización del conocimiento: una propuesta interdisciplinar. In: GARCIA MARCO, F. J. (Ed.). *Organización del conocimiento en sistemas de información y documentación*. Zaragoza: Librería General, 1995. p. 219- 228.

GIASSON, J. *A compreensão na leitura*. Lisboa: Asa, 1993. 317 p.

HJØRLAND, B. Epistemology and the socio-cognitive perspective in information science. *Journal of the American Society for Information Science and Technology*, v. 53, n. 4, p. 257-270, 2002.

LARA, M. G. de. Problemas da organização do conhecimento na contemporaneidade. In: DODEBEI, J. A. C. G. (Org.). *Complexidade e organização do conhecimento: desafios de nosso século*. Rio de Janeiro: ISKO-Brasil; Marília: FUNDEPE, 2013. p. 238-245.

PINTO MOLINA, M. *Análisis documental: fundamentos y procedimientos*. 2. ed. rev. y aum. Madrid: EUDEMA, 1993. 270 p.

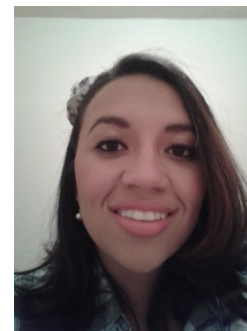
SARACEVIC, T. Information science. *Journal of the American Society for Information Science*, v. 50, n. 12, p.1051-1063, 1999.

VAN DIJK, T. A. (Ed.). *Discourse as structure and process*. London: Sage, 1997. 368 p.

Information organization, representation, retrieval and access: (re)configuration of MARC21 Format and BIBFRAME for cultural diversity in digital information environments?



Zaira Regina Zafalon
São Carlos Federal University
zaira@ufscar.br



Marcela Cristina Néspoli
São Carlos Federal University
marcela_nespoli@hotmail.com

1 Introduction

Studies in Information Organization and Representation have been redefined over the years in search of a guarantee of information retrieval and access especially due to the diversity present in cultural heritage institutions. Such diversity can be characterized by search behavior, retrieval mode and access to information, (re)configuration of informational production and adopted technologies. In this way, concepts, models, standards and formats of information organization and representation have also been discussed. This research is part of this universe and seeks to discuss the purposes of MARC21 Format and BIBFRAME as standards of information representation in digital information environments. The discussions involve, as a background, the expansion of the universe of information search and retrieval by communities as well as metadata use and reuse by technological parks; both relate to the availability of information by cultural heritage units on the web through the description of resources. In order to meet the proposed objective, the research presents a qualitative approach, of an applied nature, with a descriptive and exploratory profile. At the same time, it uses bibliographical and documentary research. Because this is a recent theme, this paper intends to serve as a social and theoretical contribution for future research on this topic.

2 Background on representation of resources

Information organization and representation (IOR) actions are based on processes of metadata analysis, synthesis, condensation, representation, retrieval, use and reuse of information resources in order to coexist with the user and obtain documents that seek to meet their needs given the possibilities of establishing communication. We highlight the studies focused on the concepts, models, standards and formats of information organization and representation, whether they are aimed at reading, analyzing, condensing, representing and constructing documentary languages, whether they are of functional requirements, conceptual modeling, interoperability and structures of information sharing and access by humans and machines.

The result of documentary representation efforts, present in IOR studies, focuses on the bibliographic record that establishes, therefore, semantic relations, between the elements present in the represented document and their value, subsemantics, marked by the relation between the several units that compose a bibliographic record and supersemantic, among the various records that refer to the various manifestations of a work, in addition to the syntactic relations, a requirement for the registration of value in semantic structures. (ZAFALON, 2014).

Representation standards, both the content and structure ones, ensure record consistency by defining syntax and semantic rules for the description of informational support and the content of informational resources, and attest to the success in resource retrieval once it minimizes possible differences in information representation and sharing.

3 (Re)configuration of purposes by cultural diversity?

MARC21 format, as a metadata structure standard, taken as a requirement for the process of sharing bibliographic records, provides mechanisms for reading and interpreting bibliographic data by machines. Allied to the records exchange format (ISO 2709), content standards (CDD, CDU, AACR, LCSH, among others) and communication protocols and bibliographic retrieval (Z39.50), ensures the transfer of qualified data across machines. Avram (1970, 1975) points out that the development of the format was based on the philosophy of information sharing and reuse, in which

a set of data can be accessed and used by several users at the same time. Initially configured for use by the Library of Congress (LC), MARC assumes features that allow each library to have autonomy both in providing records and in sharing with other units. MARC's flexibility, according to Zafalon (2008), occurs through its marks, given the adopted conventions which make the management and retrieval of bibliographic data more efficient, resulting from the specificity they promote. MARC was, therefore, developed with the aim of standardizing the representation of informational records and, thus, allow, through cooperative actions, the efficient exchange of bibliographic data in order to avoid duplication of efforts. Disadvantages of the MARC format, discussed by Tennant (2002, 2013), one of the supporters of the MARC format substitution initiative, indicate the complexity of its adoption, reliance on markup for semantic effects, record-keeping by qualified professionals, which generates desertion and resistance from the community itself, and the difficulty of reading the records given the complex syntax, natural of the numerical scheme of the labels. The author also highlights, the inability of the format to perform document relationships that are entirely digital, such as the URLs.

In this context, the Bibliographic Framework (BIBFRAME) model, from LC initiative, has been proposed as a substitute for MARC since 2012, whose purpose is to serve as a model for which several content models can be mapped. It is being designed to, intentionally, become a formal entry point for the librarian community into a network that involves both metadata and linked documents. This tends to be made possible by the fact that BIBFRAME is based on the proposal of creating semantic relations (LinkedData) and for presenting a modeling based on Resource Description Framework (RDF), model for data interchange on the Web. In the face of information reuse paradigm at its highest level, BIBFRAME recognizes entity, attributes and relationships, through its main classes (work, instance, author and annotation). (LIBRARY OF CONGRESS, 2012).

By aligning MARC's objective with the need to make bibliographic resources available to users, to the possibility of access to bibliographic resources, and, finally, to the objective of guaranteeing access to bibliographic resources, it is understood that MARC meets its initial purpose, since it guarantees retrieval and access to the bibliographic resources described in it. However, in a broader context involving open data and Web data studies, for example, MARC appears obsolete given the difficulty in creating relationships between resources, which compromises the navigation goal

beyond the catalog. BIBFRAME, however, in terms of bibliographic data structuring, in order to allow data exchange, as well as to enable the retrieval of increasingly specific information, indicates, for the moment, a gap in the new model, which has generated debates and disagreements.

4 Conclusions

MARC format has been harshly criticized for the number of tags, fields and subfields not always used which carry unnecessary costs to the software, as well as the difficulty for catalogers to use it. BIBFRAME, because of its low specificity in record description, jeopardizes the search result and the search refining, which requires fine granularity. BIBFRAME provides the description in instances and relationships with other major classes (work, author and annotation), but not the specific description. This is due to the data modeling used, RDF, which does not involve the creation of as many predicates as necessary to achieve the specificity of MARC. Another point refers to the temporality of URLs and the standardization in the creation of URIs, in order to guarantee univocity in retrieval, as foreseen in MARC. Obtaining control of authorities and access is imperative to achieve the purpose of the model.

Even with so many issues and uncertainties, LC and Zepheria (a company specialized in library applications) are developing a demo version of the model for bibliographic resource description to structure data and allow it to be entered. It is not clear, however, how data will be stored and exchanged with other formats. The difficulty is that RDF has the function of describing generic relationships across data, allowing computers to use and exchange these data and still preserve its original meaning, but not structure the information.

In order to draw a comparison between MARC21 and BIBFRAME, it is observed that the first one is dedicated to allowing the communication of bibliographic data through exchange in order to meet any informational need, anywhere worldwide, by any library or user; and that the second one seeks to integrate various knowledge communities, through linked data. As for the structure, both are differentiated since MARC uses fields and subfields, while BIBFRAME adopts RDF triples. MARC's advantages can be observed when saving time, inserting multiple records, consistency of records, guarantee of data integrity, possibility of cooperative cataloging and guarantee of more efficient retrieval. BIBFRAME, on the other hand, presents

integration of several knowledge communities, possibility to provide better navigability, self-sufficiency in records description, library entry in the Web data. The counterpoint in using BIBFRAME focuses on the lack of familiarity by users with RDF and the fact that it is not structured for information exchange, making use of little specificity in the descriptions.

When resuming the general objective of the research, it is understood that BIBFRAME arises with the intention of integrating libraries to all knowledge communities, through the Web, since, due to its structure, MARC21 format cannot carry out such characteristic. In this way, it can be verified that the substitution proposal is based on the need to integrate libraries with the Web environment in order to allow greater navigability to the user, integrating concepts and models such as FRBR, ontologies, RDF, XML and LinkedData.

In this sense, both MARC and BIBFRAME are intended for integrating communities that are or were isolated. MARC, with a somewhat timid purpose compared to BIBFRAME (disregarding the cultural and technological configuration of the time), connected the libraries so they could exchange bibliographic records, which resulted in benefits such as data consistency, information reuse and decrease in rework. BIBFRAME, through navigability concept, tends to integrate the library into a wide range of knowledge communities on a macro level. However, the similarity of purposes is restricted only to this point, since there are still many gaps and uncertainties in the development of BIBFRAME that do not contemplate the objectives served by MARC and, therefore, do not designate the same role played by the format.

Thus, it is not envisaged, at least for now, that BIBFRAME may be a substitute for MARC. However, similarly to what occurred to MARC, many of its applications and advantages as a model of data exchange could not be felt until its stability and conception as a consolidated format. The same can happen with BIBFRAME, which, as a model, does not provide a solution to all doubts that still linger.

In short, although BIBFRAME initiative has arisen for updating MARC21, it is understood that there is no impediment of coexistence between the two, since each serves specific and complementary purposes.

References

The references was made folowing the ABNT rules.

AVRAM, H. D. MARC: its history and implications. Washington: Library of Congress, 1975.

AVRAM, H. D. The evolving MARC system: the conception of a data utility. Illinois: University of Illinois, 1970.

LIBRARY OF CONGRESS. Bibliographic Framework as a Web of Data: Linked Data Model and Supporting Services. Washington, DC. 2012. Disponível em: <www.loc.gov/bibframe/pdf/marclid-report-11-21-2012.pdf>. Acesso em: 28 abr. 2015.

TENNANT, R. MARC Must Die. Library Journal, 15 Oct. 2002. Disponível em: <<http://lj.libraryjournal.com/2002/10/ljarchives/marc-must-die/>>. Acesso em: 01 out. 2014.

TENNANT, R. The Post MARC Era, Part 2: Where the problems lie, part 1. Library Journal, 8 May 2013. Disponível em: <<http://www.thedigitalshift.com/2013/05/roy-tennant-digital-libraries/the-post-marc-era-part-2/>>. Acesso em: 25 set. 2014.

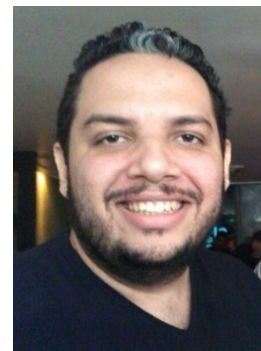
ZAFALON, Z. R. Formato MARC21 Bibliográfico: estudos e aplicações para livros, folhetos, folhas impressas e manuscritos. São Carlos: EdUFSCar, 2008.

ZAFALON, Z. R. Scan for MARC: conversão de registros bibliográficos analógicos para o Formato MARC21 Bibliográfico. São Paulo: Editora Unesp, 2014.

Music, literature and audiovisual: the contributions of knowledge organization (KO) in the intersectional relations between the works by Dorival Caymmi and Jorge Amado



Fabio Assis Pinho
Federal University of Pernambuco
fabiopinho@ufpe.br



Francisco Arrais Nascimento
Federal University of Pernambuco
francisco.arrais.nascimento@gmail.com



Andréa Carla Melo Marinho
Federal University of Rio Grande do Sul
andreacmmarinho@gmail.com

Introduction

Northeast Brazil has been evidenced in its historiographical construction by its climatic, economic and political characteristics, thus building a distorted image, allocating it as a place of drought, poverty and colonialism. It is understood that discursive echoes of such stereotype influence the production of cultural artifacts guided by control devices, thus creating a universe of social representations of the region. According to Albuquerque Jr. (2010, p.30):

The stereotypy discourse is an assertive, repetitive discourse, it is an arrogant speech, a language that leads to uncritical stability, it is the fruit of a secure and self-sufficient voice that claims the right to say what the other is in a few words. The stereotype arises from a gross and indiscriminate characterization of a strange group, in which

individual multiplicities and differences are erased in the name of superficial similarities of the group.

We can understand that cultural artifacts represent not only forms of stereotyping, but also elements of resistance to social proselytism, identity and social representations. This claim can be proved by turning to the cultural aspects of Northeast of Brazil, where social actors such as Jorge Leal Amado de Faria (1912-2001) and Dorival Caymmi (1914-2008) made their artistic productions (literature and music, respectively), credible representations of Bahian daily life, from which sectarian characters that outlined the social structure of the State of Bahia in the 20th century emerge, such as retirees, blacks, colonels, politicians, merchants, bohemians and prostitutes.

Therefore, the objective of the research is to analyze the relations between the production of Jorge Amado and Dorival Caymmi in audiovisual artifacts. For this, an intersectionality study established among the artists' productions will be used through the movie and television adaptations of Jorge Amado's literary work that had Dorival Caymmi's songs as soundtrack. According to CRENSHAW (2002, p.177):

Intersectionality is a conceptualization of the problem that seeks to capture the structural and dynamic consequences of the interaction across two or more axes of subordination. It specifically copes with the way in which racism, patriarchy, class oppression and other discriminatory systems create basic inequalities that structure the positions regarding women, races, ethnicities, classes, and others.

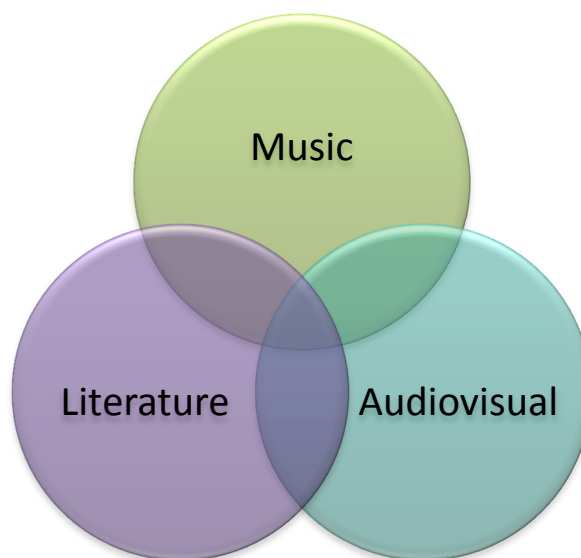


Figure 1 - Intersectionality among cultural artifacts

Source: by the authors

According to Ferreira (2001) music is understood as the art and science of combining sounds in a pleasant way to hearing, as well as written in the form of musical composition (inscribed music). The study presented here points to music as an information resource, given its documentary nature, thus focusing on the materiality of music, understood in the form of knowledge inscription (document), its descriptive and thematic elements analyzed from the perspective of theoretical and methodological assumptions of Information Organization (IO) and Knowledge Organization (KO) to understand and analyze them.

According to Ferreira (2001), literature is defined as being the art of composing artistic works in prose and verse. While a document is characterized as literary work and thematic treatment is given from the nationality of the author and literary genre.

Cultural artifacts of an audiovisual nature have their genesis in film production, which according to Moreno (1995) aim to transmit a message through a composition of sound and image, obeying codes that are established by the cinematographic technique and originated in social demands.

Methodology

Regarding the methodological procedures, the techniques of document mapping and subject analysis were used, understanding the descriptive nature of the study. Therefore, this study is a documentary research, supported by a bibliographical

research, that uses the combination of methodologies presenting a proposal for documentary structure representation.

From this methodology, we identified 13 books by Jorge Amado, 22 songs by Dorival Caymmi and 26 audiovisual productions: 15 films, 4 soap operas, 6 series, 1 miniseries and 1 special case. These productions were searched on Cinemateca Nacional (www.cinemateca.gov.br) and on Arquivo Memória Globo (memoriaglobo.globo.com).

According to Guimarães (2008), starting from the social, materialized and cyclical conception of knowledge, the focus lies on the search for understanding, organization and representation of this knowledge to make it available to a larger number of people. From this perspective, it can be understood that the content of fictional narrative works (audiovisual, literature and music) can be characterized as information that needs to be organized for better retrieval.

Results and discussion

All of Amadian and Caymmian works are linked to the protection of "popular values" through literature. They aim to maintain the traditions and practices of everyday life alive and to represent people as the legitimate source of the national identity, in general, and especially from Bahia,.

Therefore, the conception of language as a representation of the world is discussed within the scope of IS as an element of knowledge organization through Dahlberg's concept theory (1978):

Knowledge was fixed through language elements. New knowledge appears with new linguistic elements and also through these become clearer and distinct. We can say that this process of growth will last as long as man exists on earth and use language as the expression of his thoughts.

The use of language as an expression of human thought is manifested through social relations and the communication processes between the subjects and their peers. This manifestation can also develop through artistic-cultural phenomena, which materialize in cultural artifacts (cinema, dance, sculpture, literature, music, painting and theater). These objects are informational resources for the preservation and protection of the memory of social events.

Information treatment involves two specialties identified as: descriptive treatment and thematic treatment. The first aims to identify the descriptive elements of a document, such as: author, title, year of publication, etc. The second aims to identify the document from its thematic content, that is, the subject.

The process of identifying the document subject involves the extraction of concepts and consequently points to the construction of models of information representation. Hjørland (1992), when discussing the concept of subject characterized their forms of treatment and presented a methodology constituted from the investigation of specific knowledge domains through the identification and contextualization of their cultural, historical and linguistic conditions.

Tennis (2012) presents the definitions of scope and reach of a domain analysis, as well as its purpose. Regarding the scope and reach, he states that it is necessary to specify its extension, naming the domain and specifying its extensions and exclusions. The purpose of the analysis is related to its genre (descriptive or instrumental). Therefore, according to the author's proposal, the research presents a scope in the audiovisual productions from the adaptations of the literary works of Jorge Amado that present compositions by Caymmi in their soundtrack with a descriptive purpose and developed through subject analysis techniques.

Thus, it is possible to perceive that the materiality of this information is constituted from a social phenomenon, that is, information passes from the Intangible state (information as knowledge) to the tangible state (information as a thing), as Buckland (1991) points out.

From this perspective, the research emphasizes the subject analysis technique of audiovisual productions from the relation of literary and musical works. The instruments of information representation (IR) are based on the study by Novellino (1996) that points out the subject analysis of a document and its result in a linguistic expression, in addition to the attribution of concepts to the analyzed document as the main steps in the IR process.

Tables 1 and 2 present the extracted elements of the audiovisual production based on Jorge Amado's work, which appear as elements of descriptive representation of the analyzed documents.

<insert tables 1 and 2 here>

Conclusion

Finally, we understand that intersectionality is a theoretical-practical concept applicable to Information Science, in which, in the analyzed case, the set of information is configured as an element of record keep of a social phenomenon revealing characteristics of a group through thematic treatment of information (Guimarães, 2008). In order to do so, it is emphasized that the objective of this research was reached and within the scope of KO, this study serves as a subsidy and points out possibilities in the field of culture and social representations.

References

The references was made following ABNT rules.

ALBUQUERQUE JR, Dúval Muniz de. **A invenção do Nordeste e outras artes**. 4^a ed. São Paulo: Cortez, 2009.

BUCKLAND, Michael K. Information as thing. **Journal of the American Society for Information Science**, v. 45, n. 5, p. 351-360, 1991.

CRENSHAW, K. Documento para o encontro de especialistas em aspectos da discriminação racial relativos ao gênero. **Estudos Feministas**, Florianópolis, v. 10, n. 1, p. 171-188, jan. 2002.

DAHLBERG, Ingetraut. Teoria do Conceito. **Revista Ciência da Informação**, Rio de Janeiro, v.7, n.2, p.101-107, 1978.

FERREIRA, Aurélio Buarque de Holanda. **Miniaurélio Século XXI: o minidicionário da língua portuguesa**. 5.ed. Rio de Janeiro: Nova Fronteira, 2001.

FOUCAULT, Michel. **Microfísica do poder**. Rio de Janeiro: Graal, 2000.

GUIMARÃES, José Augusto Chaves. A dimensão teórica do tratamento temático da informação e suas interlocuções com o universo científico da International Society for Knowledge Organization (ISKO). **Revista Ibero-americana de Ciência da Informação**, v. 1, n. 1, p. 77-99, jan./jun. 2008.

HJORLAND, Birger. The concept of "subject" in Information Science. **Journal of Documentation**, v.48, n.2, p172-200, June 1992.

MORENO, Antônio do Nascimento. **A personagem homossexual no cinema brasileiro**. 1995. 148 f. Dissertação (Mestrado em Artes) – Instituto de Artes, Universidade Estadual de Campinas, Campinas, 1995.

NOVELLINO, Maria Salet Ferreira. Instrumentos e metodologias de representação da informação. **Informação & Informação**, Londrina, v.1, n.2, p.37-45, jul/dez. 1996.

ROLNIK, Suely. **Cartografia Sentimental: Transformações contemporâneas do Desejo**. São Paulo: Estação Liberdade, 1989.

TENNIS, Joseph T. Com o que uma análise de domínio se parece no tocante a sua forma, função e gênero? **Brazilian Journal of Information Science**, Marília (SP), v. 6, n. 1, p. 3-15, jan./jun. 2012. Disponível em: <http://www2.marilia.Unesp.br/revistas/index.php/bjis/index>. Acesso em: 30 jul 2013.

Mediation in Knowledge Organization Domain



Mona Cleide Quirino da Silva Farias
São Paulo State University
monaquirino@gmail.com



Carlos Cândido de Almeida
São Paulo State University
carlosalmeida@marilia.unesp.br



Daniel Martínez-Ávila
São Paulo State University
dmartinezavila@marilia.unesp.br

Introduction

Knowledge organization in its sociocultural approaches, notably in Beitudo (1986, 1995, 2001, 2002, 2005), Campbell (2000, 2010), Dal' Evedove (2014), García Gutiérrez (1998, 2002a, 2002b, 2002c, 2004, 2006, 2008, 2011), Guimarães (2005), Guimarães and Milani (2010, 2011), Guimarães and Pinho (2008), Hjørland (2003, 2007, 2008, 2010, 2013), Hudon (1997, 1999, 2003), Milani (2010), Milani et al. (2009) and Pinho (2006, 2010) among others, impels us to reflect on the context in knowledge organization and representation. This context, after all, is constituted by interpretative and mediation processes.

It is worth mentioning that the term "mediation" is still polysemic not only in Information Science. Bortolin (2010) states that there is a "plurality" of definitions of mediation, so it does not present a consensual nor scientific concept that provides a basis for broad and coherent reflection for the various areas and lines of Information

Science. For Malheiro and Ribeiro (2011, p.156) IS lacks the conceptual use of mediation, understood as an articulating instance between different parties always in certain situations and contexts.

This article aims to reintroduce the theme of mediation as a necessary element to understand knowledge organization and representation processes through the contribution of conceptual, cultural and professional dimensions. Through an analysis of the critical literature of knowledge organization, we analyze its status as a field of mediations in which organization and representation processes can be recognized as social mediation, whose activities consequently promote the mediation of knowledge among subjects, considering their values and goals capable of recognizing other points of view and the different communities, values and cultures.

Mediation in Knowledge Organization

The Conceptual Dimension

Mediation should be seen as an interpretative social process, and as such requires activities that aim at knowledge organization, i.e. that can cover ethical and sociocultural aspects for knowledge treatment and socialization, insofar as considers the multiple contexts and goals of the communities.

Mediation corresponds to an action that tries to learn about the contexts, cultures, subjects and communities, in order to reflect and rethink the modes of knowledge, as knowledge is constantly modeled. In this aspect, we understand that knowledge organization is a mediating field, where knowing the contexts of discursive communities is a necessity.

Mediation as an interpretive process of social and cultural order must be analyzed from the context in which it proposes to know and interpret. It is through the context that one can get to know cultures and communities. In this sense, the human figure is justified, because it brings all its referential of knowledge, beliefs, ethics, languages, which influences the action of mediating in the process of knowledge representation.

For this purpose, mediation should also be analyzed as an instrument for transformation of subjects. The professional is the interlocutor of such process, considering his contribution evidenced through knowledge organization and

representation. Knowledge organization is thus considered a field whose processes are knowledge mediators. This militant element of mediation practices as it is present in Freire (1976), Almeida Júnior (2003, 2007, 2009) and Bortolin (2010), although questionable, understanding its ethical commitment and social responsibility is fundamental.

Thus, mediation is a dialectical process, since the relations between subjects and the numerous forms of knowledge do not consist of something reified. However, to represent knowledge, whatever it is, one must take into account ideas, worldviews, contexts, domains of knowledge, experiences, languages and the very notion of culture.

We understand that every form of knowledge organization is a mediation between knowledge and user communities. The user idea makes little sense when we decide to assume the social role of mediation in knowledge representation, because after all, we deal with communities, with historical subjects. A classification is the result of mediation, just like the thesaurus. They are not mere tools or products of knowledge organization, but emanate from the social praxis of mediation as a properly human capacity. For this, knowledge organization implies a continuous process of dialogue.

Mediation is a communicative and dialogic process, so performed action is not a neutral intervention to an external reality. The same goes for communities of knowledge organization and representation systems. The functionalist meaning of "system" also falls into disuse with the compromised conception of mediation. In the search for informational contents, communities bring with them values, categories, representations, histories, languages, that is, culture in the anthropological sense. A knowledge domain does not supplant these variables.

The knowledge organization, whether we accept it or not, is a field of mediation. We take the idea of field by Bourdieu (1983, 1989, 2004), in the sense that it is understood as a space of conflicts, of struggles for positions, of disputes between forms of representation of social reality. The results of knowledge organization do not operate on emptiness. On the contrary, they imply activities of interpretive analysis that interfere and, consequently, represent a discursive community. Knowledge representation is mediation that, according to Fernández-Molina et al. (2005), can be understood as an activity that seeks to provide information to several users - i.e., communities - without distinction and with the purpose to make users feel reflected in the represented context.

Under this idea of mediation, the postmodern and sociocultural trends in knowledge organization, in various shades, would be brought together as the materialization of the mediation process carried out in information spaces.

The Cultural Dimension

Mediation takes place in a sociocultural reality, which must be understood with greater detail. A simple enumeration of measurable variables would not respond to the knowledge of the context. The assessment of its behavior, considering its episodic function of user of an information retrieval system, does not represent the social dimension either. A demographic analysis of the community, although necessary, does not exhaust the knowledge of the context, much less substitutes it.

If we assume mediation as a founding instance, the context to which it refers is notably multicultural: "The process of knowledge representation, as well as the instruments that permeate it and the products that flow from it, are not neutral, imbued with moral values" (FERNANDEZ-MOLINA et al., 2005, p 204).

Mediation in this sense is a social action, an intervention, in the face of multicultural contexts. Multiculturalism, according to Fernández-Molina et al. (2005), brings to society the need to recognize and respect differences, in order to promote the coexistence of respect among individuals and groups.

In this sense, García Gutiérrez (2002a) proposes the substitution of the information professional by the mediator, considering that this denomination is more comprehensive, due to the mediator's performance in multicultural contexts.

In order to do so, multicultural mediation takes on an effective complexity as it must devote attention both to the moment of knowledge reception and to the possibilities of its fragmentation, considering that one of the multiculturalism challenges is to try to avoid incommunicability situations and thus seek new mediation and translation spaces (FERNÁNDEZ-MOLINA et al., 2005).

Guimarães and Pinho (2008) affirm that cultural universe cannot be reduced to a positivist paradigm to the extent of not noticing reality, allowing the prevalence of dominant positions. The authors also warn that knowledge organization and representation systems need to have, in their foundations, ethical principles that promote cross-cultural view and allow the dissemination of various aspects related to knowledge.

The mediator must try to understand the multiple contexts and their languages, as well as their various forms of expression represented by the subjects of a socio-cultural space of knowledge. In this aspect, the professional must be attentive to the linguistic limitations in the search to avoid "linguistic imperialism", emphasized by Hudon (1997).

In this discussion, the author highlights multilingual thesauri and considers that, in order to face language barriers, multilingual thesauri can act in a way that allows users to access information considering different cultures and languages.

It is worth emphasizing that mediation occurs through the expression of languages used, for example, in a certain representation system elaborated by professionals. Thus, the construction of knowledge schemes should reinforce the guarantee of multicultural issues that should be highlighted by professionals at the moment of mediation, in representation. The issues related to multiculturalism reflect ethics, in what corresponds to the defense of values and inclusion in order to safeguard the citizens' rights (FERNÁNDEZ-MOLINA et al., 2005).

Culture is a complex of signs that deserves attention when elaborating knowledge organization and representation, especially because it is linked to language and, consequently, requires ethical and social care. According to García Gutiérrez (2004, 83), "Culture, considered as symbolic and intangible imaginary of the imaginary, is a phenomenology exclusively conceived as an open system, that is, to think of it as a closed and formal system is a mistake".

According to this perspective, mediating professionals must be willing to perceive that in the field of culture there is a diversity of codes, symbologies and languages that can assist them in the dialogue process with communities. Such dialogue establishes the relationship between professionals and subjects, subjects and cultural contexts, professionals and representation systems, knowledge organization and communities.

Multiculturalism is not only a perspective, an exoteric theoretical approach, devoid of practice, but it is social reality itself, the condition of our societies. We live together, as Garcia Canclini (2001) argues, in the same reality, in the same space with multiple historical temporalities. If we cross Brazil, we find millennial cultures coexisting with European simulacra of society, regarding the material conditions of life. These cultures, which represent most part of the more than 200 languages spoken in Brazil, are not adequately represented by knowledge organization systems. Similarly, if we

move to European countries, we note the complexity of language policy planning, since in addition to the 24 official languages of the European Union, there are many other linguistic groups not included. Consequently, multiculturalism cannot be used only as a discourse by knowledge organization, because, behind it, there is a hybrid cultural reality to which we must undertake actions mediation.

Thus, it is imperative to consider cultural diversity, such as the context mentioned by Fernández-Molina et al. (2005), which recognizes the need to explicit values and paradigms that permeate historical, social and cultural aspects in the knowledge organization and representation.

However, there is an ethical issue involved in the action of mediation. Guimarães (2000) categorizes five instances corresponding to the ethical attributions that involve professional performance: the instance related to users, to the organization, to the information, to the profession and to the professionals themselves. In the scope of this discussion, such instances reinforce mediation as an ethical commitment to community. Thus, we make room for the professional consequences of mediation.

The Professional Dimension

Mediation, in the field of knowledge organization, is shaped, in practice, as a fundamentally interpretative process, which implies ethical and dialogic action on knowledge as the basis is supported on contexts, languages and understanding of culture. We understand that the sociocultural studies referred above envisage a professional practice, whether performed by librarians, archivists, museologists, etc., as mediators of knowledge organization and representation.

Regarding professional practice, Guimarães (2005) highlights that the "professional dimension of ethics as a set of values that a social segment, characterized by the specificity of knowledge and making (profession) establishes as necessary and fundamental to the exercise of that profession" .

According to Fernández-Molina and Guimarães (2002), in the core of discussions addressed in Information Science, in the thematic of the professional have been treated with greater emphasis the discussions related to professional practice, or an approach on the problems of specific rights, that is, to specificity, copyright, right to intellectual freedom, censorship, among others.

In addition, we highlight the existence of some works that emphasize the mediator's ethical role in knowledge organization, such as, in media context (GARCÍA GUTIÉRREZ, 2013, 2014, GARCÍA GUTIÉRREZ, MARTÍNEZ-ÁVILA, 2014a). and also to critical formation of the professional in knowledge organization (GARCÍA GUTIÉRREZ; MARTÍNEZ-ÁVILA, 2014b).

According to Fernández-Molina et al. (2005), information professionals have the responsibility to ensure the diversity of access to cultural information. In this sense, their work concerns the projection, assessment, maintenance and review of representation systems. Thus, mediation is a sociocultural process of hybrid character related to the practice of professionals in knowledge organization.

The professional is a social being that reflects, dialogues and interacts with the multiple contexts of the various communities. In the scope of processes, as well as in the activities of knowledge organization, Guimarães and Pinho (2008) consider that representation implies an intellectual activity and demands a conscious and critical attitude from professional, without ignoring knowledge on historical and social aspects surrounding registered and socialized knowledge.

We understand that, within the scope of the knowledge organization, interests are already evident with professional and training issues of personnel to act as mediators. In this same line, we must highlight the work that reconfigured the vision of techniques, products and instruments of knowledge organization, which now, under a critical and non-neutral view, motivate professionals to work for a society that recognizes the multicultural fact and the semiotic role they represent in knowledge organization.

The professional dimension of mediation in knowledge organization is based on the ethical commitment of the professional, from content representation and construction of a classification system to the recognition of the need to learn sociological, anthropological and linguistic methods to provide integrated mediation.

Conclusion

In short, we rely on the words of García-Gutiérrez and Martínez-Ávila (2014a), as they assume the theoretical recognition of the subjective presence and biases of knowledge organization as mediation, a work that was generally considered neutral, banal and even objective and can be applied in practice through a critical training of professionals. As a conclusion, we should briefly list the arguments made so far.

Firstly, recognition of non-neutrality by instruments and professionals who manage them is important, as Hjørland (2003), Mai (2004) and Olson (2002) point out. In this way, it is necessary to say that values are contained in representational instruments such as thesauri, faceted classification systems, documentary languages, and in some cases, the positions of the professionals involved.

Secondly, knowledge organization must be a critical and interpretative space that favors the practice of mediation, in view of an ethical, multicultural and adequate performance to the characteristics of the diverse communities that make use of this knowledge.

A third argument is that, as Garcia Gutiérrez (2002a) suggests, the idea of an information professional or specialist in knowledge organization systems - which seems to us the description of a function or task - , performed by the mediator, considering the scope of the action in multicultural contexts.

Finally, knowledge organization is not a type of mediation, or a specific case in the information professionals' activities, it is rather equivalent to mediation. In this sense, it would be fair to consider that we do work in knowledge mediation, not only in knowledge instrumental organization.

References

The references was made following ABNT rules.

ALMEIDA JÚNIOR, O. F. Mediação da informação: discutindo a atuação do bibliotecário. In: FADEL, Barbara (Org.). **A informação nas organizações sociais: desafios em face de multiplicidade de enfoques**. Marília: FUNDEPE, 2003. 1 CD-ROM.

_____. Mediación e información. **Ibersid**, p. 27-35, 2007.

_____. Mediação da informação e múltiplas linguagens. **Tendências da Pesquisa Brasileira em Ciência da Informação**, v. 2, n. 1, p. 89-103, jan./dez. 2009. Disponível em: <<http://inseer.ibict.br/ancib/index.php/tpbci/article/view/17/39>>. Acesso em: 21 de julho de 2015.

BEGHTOL, C. Semantic Validity: concepts of warrant in bibliographic classification systems. **Library Resources & Technical Services**. v. 30, n. 2, p. 109-125, 1986.

BEGHTOL, C. Domain analysis, literary warrant, and consensus: the case of fiction studies, **Journal of the American Society for Information Science and Technology**, v. 46, n. 1, p. 30-44, 1995.

BEGHTOL, C. Relationships in Classificatory structure and meaning. In: C.A. Bean & R. Green, (Eds.). **Relationships in the organization knowledge**. Dordrecht, Netherlands: Kluwer, 2001, p. 99-113.

BEGHTOL, C. Universal concepts, cultural warrant, and cultural hospitality. In: LÓPEZ-HUERTAS, M. J. (Ed.). **Challenges in knowledge representation and organization for the 21 century**: integration of knowledge across boundaries. Würzburg: Ergon-Verlag. 2002. p. 45-49. (Advances in Knowledge Organization, 8).

BEGHTOL, C. Ethical decision-making for knowledge representation and organization systems for global use. **Journal of the American Society for Information Science and Technology**, v. 56, n. 9, p. 903-912, 2005.

BORTOLIN, Sueli. **Mediação oral da literatura**: a voz do bibliotecário lendo ou narrando. 2010. 232f. Tese (Doutorado em Ciência da Informação)- Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2010.

BOURDIEU, P. O campo científico. In: ORTIZ, R. (Org.). **Pierre Bourdieu**: sociologia. São Paulo: Ática, 1983.

BOURDIEU, P. **O poder simbólico**. Lisboa: DIFEL; Rio de Janeiro: Bertrand Brasil, 1989.

BOURDIEU, P. **Os usos sociais da ciência**: por uma sociologia clínica do campo científico. São Paulo: Unesp, 2004.

CAMPBELL, G. Queer theory and the creation of contextual subject access tools for gay and lesbian communities. **Knowledge Organization**, v.27, n. 3, p. 122-131. 2000.

CAMPBELL, G. Tensions between language and discourse in North American knowledge organization, **Knowledge Organization**, v. 37, n. 1, p. 51-57. 2010.

DAL' EVEDOVE, P. R. **O tratamento temático da informação em abordagem sociocultural**: diretrizes para definição de política de indexação em bibliotecas universitárias. 2014. 259f. Tese (Doutorado em Ciência da Informação) Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2014.

FERNÁNDEZ-MOLINA, J. C.; GUIMARÃES, J. A. C. Ethical aspects of knowledge organization and representation in the digital environment: their articulation in professional codes of ethics. In: LÓPEZ-HUERTAS, M. J. (Ed.). **Challenges in knowledge representation and organization for the 21 st century**: integration of knowledge across boundaries. Würzburg: Ergon-Verlag, 2002. p. 487-492. (Advances in Knowledge Organization, 8).

FERNÁNDEZ-MOLINA, J. C.; GUIMARÃES, J. A. C.; VIDOTTI, S. A. B. G.; FLAMINO, A. N.; SOUZA, A. S.; CAMARGO, L. S. A.; SILVA, M. S.; MORENO, P. S.; RAMALHO, R. A. S. Aspectos éticos de las nuevas tecnologías de la información y la comunicación y su reflejo en la organización del conocimiento. In: GASCÓN, J.; BURGUILLOS, F.; PONS, A. (Org.). **La dimensión humana de la organización del conocimiento**. Barcelona: Universitat de Barcelona, 2005. p. 177-186.

FREIRE, P. **Ação cultural para a liberdade e outros escritos**. São Paulo: Paz e Terra, 1976.

GARCÍA CANCLINI, N. **Culturas híbridas**: estrategias para entrar y salir de la modernidad. Buenos Aires: Paidós, 2001.

GARCÍA GUTIÉRREZ, A. **Principios de lenguaje epistemográfico**: la representación del conocimiento sobre Patrimonio Histórico Andaluz. Sevilla: Instituto Andaluz del Patrimonio Histórico, 1998. (Cuadernos técnicos, 3).

_____. Knowledge organization from a “culture of the border”: towards a transcultural ethics of mediation. In: LÓPEZ-HUERTAS, M.J. (Ed.). **Challenges in knowledge representation and organization for the 21st century**: integration of knowledge across boundaries. Würzburg: Ergon-Verlag, 2002a., p. 516-522. (Advances in Knowledge Organization, 8).

_____. **La memoria subrogada**: mediación, cultura y conciencia en la red digital. Granada: Editorial de la Universidad de Granada, 2000b.

_____. Exomemoria y cultura de frontera: hacia una ética transcultural de la mediación. In: **VII Congreso Internacional sobre Organización del Conocimiento celebrado en Granada** (España), 2002c.

_____. **Otra memoria es posible**: estrategias descolonizadoras del archivo mundial. Sevilla: Universidad de Sevilla, 2004.

_____. Cientificamente favelados: uma visão crítica do conhecimento a partir da epistemografia. **Transinformação**, Campinas, v.18, n.2, p.103-112, maio/ago. 2006.

_____. **Outra memória é possível**: estratégias descolonizadoras do arquivo mundial. Petrópolis: Vozes, 2008.

_____. Desclassification in knowledge organization: a post-epistemological essay. **Transinformação**, Campinas, v.23, n. 1, p. 5-14, jan./abr., 2011.

_____. Entrevistas periodísticas y textos declarativos: un procedimiento de análisis y recuperación documentales basado en modos de búsqueda pronominal. **El profesional de la información**, v. 22, n. 4, p. 315-325, 2013.

_____. Análisis documental de noticias de prensa en sistemas de información factual. **Revista española de documentación científica**, v. 37, n. 2, e046, 2014.

GARCÍA-GUTIÉRREZ, A.; MARTÍNEZ-ÁVILA, D. Critical Organization of Knowledge in Mass Media Information Systems. **Knowledge Organization**, v. 41, n. 3, p. 205-216, 2014a.

GARCÍA-GUTIÉRREZ, A.; MARTÍNEZ-ÁVILA, D. Critical Organization of Knowledge in Mass Media Information Systems. **El profesional de la información**, v. 23, n. 5, p. 493-500, 2014b.

GUIMARÃES, J. A. C. O profissional da informação sob o prisma de sua formação. In: VALENTIM, M. L. P. (Ed.). **Profissionais da informação**: formação e atuação profissional. São Paulo: Polis, 2000. p. 53-70.

_____. Aspectos éticos em organização e representação do conhecimento (O.R.C): uma reflexão preliminar. Conferência apresentada na mesa- redonda: Organização do conhecimento e interdisciplinaridade. In: **Seminário: Memória, informação e organização do conhecimento**: cruzando fronteiras da identidade. Rio de Janeiro, Unirio, 25 e 26 de agosto de 2005. Marília, 2005. 24p.

GUIMARÃES, J. A. C.; MILANI, S. O. Bias in the indexing languages: theoretical approaches about feminine issues. En: GNOLI, C.; MAZZOCCHI, F. (Eds.). **Paradigms and conceptual systems in knowledge organization**. Italy: Ergon Verlag. 2010. p. 424- 429. (Advances in Knowledge Organization, 12).

_____. O. Problemas éticos em representação do conhecimento: uma abordagem teórica. **DataGramZero**, v.12, n.1, fev./2011.

GUIMARÃES, J. A. C.; PINHO, F. A. Reflexiones acerca de las teorías sobre ética en la representación del conocimiento. **Biblios**. Revista de bibliotecología y Ciencias de la Información . Jul./set., p. 1-12, 2008.

HJØRLAND, B. Fundaments of knowledge organization. **Knowledge Organization**, v.30, n. 2, p.87-111, 2003.

_____. Semantics and Knowledge organization. **Annual Review of Information Science and Technology**, v.41, n.1, p. 367-405. 2007.

_____. What is Knowledge organization (KO)?. **Knowledge Organization**, v. 35, n.160 2/3 p. 86-101. 2008.

_____. Concepts, paradigms and knowledge organization. In: GNOLI, C.; MAZZOCCHI, F. (Eds.). **Paradigms and conceptual systems in knowledge organization**. Italy: Ergon Verlag. 2010. p. 38-42. (Advances in Knowledge Organization, 12).

HJØRLAND, B. Theories of knowledge organization-Theories of knowledge. **Knowledge Organization**, v. 40, n. 3, p 169-181. 2013.

HUDON, M. Multilingual thesaurus construction: integrating the views of different cultures in one gateway to knowledge and concepts. **Knowledge Organization**. v. 24, n.2, p.84-91. 1997.

_____. Accessing documents and information in a world without frontiers. **The Indexer**, London, v. 21, n. 4, p. 156-159, 1999.

_____. True and tested products: thesauri on the web. **The Indexer**, London, v. 23, n. 3, p. 115-119, 2003.

MAI, J.-E. Classification in context: relativity, reality, and representation. **Knowledge Organization**. v.31, n. 1, p. 39-48. 2004.

MALHEIRO, A.; RIBEIRO, F. **Paradigmas, serviços e mediação em Ciência da Informação**. Recife: Nectar, 2011.

MILANI, S. O. **Estudos éticos em representação do conhecimento**: uma análise da questão feminina em linguagens documentais brasileiras. 2010. 141 f. Dissertação (Mestrado

em Ciência da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2010.

MILANI, S.; GUIMARÃES, J. A. C.; PINHO, F. A.; MOLINA, C. F. Os desvios na representação do conhecimento em um contexto multicultural: abordagens teóricas. En: LLORET ROMERO, N. (Ed.) **Nuevas perspectivas para la difusión y organización del conocimiento**: actas del IX Congreso ISKO España. Valencia: Universidad Politécnica de Valencia, Servicio de Publicaciones, 2009. p. 204-218.

OLSON, H. A. **The power to name**: locating the limits of subject representation in libraries. Dordrecht: Kluwer Academic Publisher, 2002.

PINHO, F. A. **Aspectos éticos em representação do conhecimento**: em busca do diálogo entre Antonio García Gutiérrez, Michèle Hudon e Clare Beghtol. 2006. 123 f. Dissertação (Mestrado em Ciência da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2006.

_____. **Aspectos éticos em representação do conhecimento em temáticas relativas à homossexualidade masculina**: uma análise da precisão em linguagens de indexação brasileiras. 2010. 149 f. Tese (Doutorado em Ciência da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2010.

The value of information and language in consumer society



Luciana de Souza Gracioso
São Carlos Federal University
luciana@ufscar.br

1 Introduction

Without many geographical and cultural restrictions, we live in a consumer society and, currently, between the desire for consumption and consumption itself, there is the intermediation of information technologies. The desires of product consumption, information services, are stimulated in virtual networks, as well as mirroring desires, identity building, and desires of belonging to groups. In this environment, we have asked ourselves about the positions of Information Science and the tools of knowledge organization, aiming to consider orientations that promote significant information retrieval processes. Thus, this paper is linked to the theme that copes with social and political dimension of knowledge organization.

In this initial moment, it is necessary to highlight the relevance of the works by B. Hjørland, J. T. Tennis, M. J. López Huertas, M. Hudon, W. Mustafá El Hadi, J. D. Campbell, H. Olson, C. Begthol, A. Garcia Gutierrez, among others, who dedicate to social issues of knowledge organization. This scientific status conferred on these authors was mapped and presented by Guimarães, in a recent conference given at the Colloquium: Organization of knowledge in time: past, present and future, organizer by PPGCI (IBICT / UFRJ) in 2015. In the Latin American panorama we have noted some authors who focus on studies on social uses of language considering their implications for knowledge organization, among them: S. P. Mostafa, M. A. R. Rojas, M. N. G. Gomez, G. S. Saldanha, M. S. Novelino, E. G. Del Orrico. (SALDANHA, GRACIOSO, 2014).

Regarding the studies that involve the discussion on the value of information consumption, we have research that transits between Information Market, announced by Aldo Barreto 15 years ago to the most recent, developed by Moura and Gomes (2014) on trade and consumption in times of digital connections.

For the argumentative course that we have proposed for this paper, we follow a theoretical path based on sociology and philosophy of pragmatics, which we intend to be able to contribute to the discussions about knowledge organization in its social and political sphere in the briefly announced context.

2 Information Consumption or use?

Current search engines advocate and guide a bilateral search and retrieval practice of contents. Thanks to syntactic and semantic resources of the Web, we retrieve. But we are already hostages of the bubble filter effect (concept attributed to extreme personalization of online content according to users' search profile). The fourth law of librarianship presented by S. Ranganathan, *Save the time of the reader* (1931) seems to have been distorted by the early SEO (Search Engine Optimization) developers, although more recently they have dedicated to produce a Web with more content. The actions of searching information on the Web are monitored every second, and the use of search words and expressions has never been so closely watched by website managers. Our information user profile has been developed for a profile of information consumers.

This flow of consumption has distorted the very social organization and construction of identities. Z. Bauman, more punctually in his book *Liquid Modernity*, 2001, has helped us to reflect about this. The reading of Z. Bauman has been made and appropriated from different perspectives in Information Science. Bennett and Vavakis (2001) discuss "[...] the challenges of information and communication technologies from the perspective of knowledge management in network society [...]". Barreto, in 2008, reviews the work: *Community - Seeking Safety in an Insecure World*, written by Bauman, who analyzes the current models of community building. Based on *Liquid Modernity*, Silva and Lopes discuss the *disintermediation of information* and present the problematizations that arise for Librarianship about the use of regulation methods and content representation. With a similar approach, Pedro Demo also analyzes the ambivalences of information society in 2000. Ziller and Moura, in 2010,

discuss Semiosis and information flows: collective assemblages and the user condition in digital environments and Mantovani and Moura, ponder about Information, interaction and mobility in 2012. Albagli in 2011 analyzes the *issue of the other: from territory to digital* and presents an interesting discussion on the interests of the nation-state in creating the illusion about freedom to create identities. This author also identifies in Habermas, in his work *Inclusion of the Other* (2002), the perspective of the thinker in prioritizing the place of the citizen, conscious and aware of his rights and duties, in society. Albagli, in addressing Habermas's work (2002), identifies that the author fears that with the nation-state fragilization, an anonymously interconnected "network logic" would be enforced, supposedly regulated by the invisible hand of the global economy, in which multinational corporations are the model of conduct, and in which there would be the absence of "a universe of intersubjectively shared meanings. (HABERMAS, 2002, p.144).

Maia (2001) is another author who turns to this spectrum and signals that, before thinking about the operationalization of models or systems that enhance virtualized communication, one must think about how to encourage them coherently in the real plan. "In order to strengthen democracy, not only efficient communication structures or institutions conducive to participation are necessary, but also the right motivation, interest and willingness of citizens themselves to engage in debates" (MAIA, 2001). With this argument by Maia, we proceeded to concatenate our research proposal of problematizing how information systems and strategies of knowledge organization on the Web could be configured to intervene in the processes of information search and retrieval in order to promote greater meaning to this process. To support our final proposition, we will briefly cross the path of philosophy of pragmatics, with emphasis on the Theory of Speech Acts by J. Austin.

3 Philosophy of Pragmatics: from the context of use to speech acts in information systems

Philosophy of language contemplates a broad and open set of studies. We highlight, in this research, the philosophy of pragmatics, or pragmatism. This philosophical current had its origins in both the United States (F.C. S. Schiller, W. James, C. Peirce, J. Dewey, G. H. Mead) and in England, more specifically linked to the School of Oxford (J. Austin, G. Ryle, P Strawson, P. Grice, J. Searle). The thoughts

of L. Wittgenstein, in his second philosophical moment, would have propelled this European movement. In the distinctions of Marcondes (2006) there are two lines for the development of English pragmatics, under which we rely on this study: the contextualist conception (L. Wittgenstein, 1889-1951) and the performative conception (J. Austin, 1911-1960).

For Austin, to say is to do. "The determination of meaning can only be made from the consideration of the act that is being performed when these expressions are uttered and from the rules that make the realization of these acts possible" (MARCONDES, 2006, 220). It is the performative conception of language in which "language is basically a form of realization." According to the ethnographers of communication (John J. Gumperz and Delleys Hymes), speech acts constitute the basic unit of verbal communication, hence, a central element to human behavior that will guide their way of acting in the world.

Austin elaborated what he named the Theory of Speech Acts in order to systematize the pragmatic phenomena related to language. For the author, the speech acts aim to produce success in communication. Once again, the understanding that language is not only used in a constative way (to describe facts) but in a performative way (to accomplish something) is reinforced. According to his theory, the speech act is the basic unit of signification and could be differentiated into locutionary acts (words and sentences of a language), illocutionary (force of speech act) and perlocutionary (consequences of speech acts). Because of this, a speech act is a consequence of the relation between the speaker's communicative intention (psychological and subjective) and the social conventions in which it is inserted (with different degrees of formalities). With respect to the different degrees of social conventions, the author considers that the rules of constitution and functioning are implicit, normative, established by previously established social conduct and that, on the face of it, delimiting and analyzing all the rules involved in these conventions is impossible. In this sense, Austin proposes that analyzes of social conventions should focus on the failures of using the implicit rules. The core of his theory would be in the illocutionary forces that impel a communication action and define them as veridical, exercising, compromising, behavioral and expository forces. (GRACIOSO, 2008).

From the exposed, we understand that speech acts move the real, construct actions. The consumer market has already addressed this and massively uses language resources, mediated by virtual technologies, to promote purchasing and

consumption actions. We might then envision some intervention by Information Science in the informational pragmatics of the Web, from guiding the use of language in communication actions, since these actions materialize in life practices. This would reinforce the position of Habermas who defends the propaedeutic function of the theory of communicative action as a theory of social action, which, in turn, highlights the role of the use of language as constitutive of communicative action and this one as constitutive of social actions.

4 Results and Conclusions

The convergence between language use actions and signification related to practical actions of everyday life can help us to understand the success on the use of the Web in the onslaughts of capitalism to encourage consumption. The Internet, in general, presents itself as a neutral territory, which can serve as an instrument for attack and defense, consensus and disagreement, forgetfulness and action in the world of life. Recognizing this potential of the Web and with strategies that use imperative illocutionary acts to guide desires and individualities, the market has been able to direct consumption actions in the most elementary daily routines through the use of language in the network. Considering the scenario presented by Bauman, on the modern dilution of the State model, the model of institutions and even the model of affective relations, and recognizing the gradual immersion of society in the use of the Internet as the main channel of expression, communication and establishment of relations that permeate from affective relations to consumer relations, we seek to understand and intervene in this territory in the sense of reflecting on the use of some information resources that allow this society to add value and meaning to their communication actions, both those that use technologies and the ones that do not, because all of them have, in a greater or lesser degree, direct relation with the daily practices of life.

In a more applied plan, we understand that information professionals could assume, as part of their practices, along with the developers of information search and retrieval systems, as well as could guide all systematization, organization and representation of content already normalized at other times, and also guide the modeling of interactive software that promote and stimulate, even before interaction between users and information agents, the questioning about the contents sought.

Existing computational logics are already able to map and even anticipate "what" users have been looking for in information systems, but we have to provide, in our systems, room for these same users to question "Why?" of their searches and interaction in the network, stimulating them to answer the same question, about the content they will receive as feedback from their actions of research and interaction on the Web.

In agreement with the theories of J. Habermas and L. Wittgenstein, we consider that the validation of information actions (communication, interaction) that occur in our daily life need to extend the level of arguments to the point of being applicable to the field of life. Thus, if we can better understand the deployments of speech acts, as discriminated by Austin, to guide aggregation of value to illocutionary acts we express on the Web (those that occur in digital information exchanges in communication processes mediated by technology and which imperatively command us to do things), we can also intervene in the practical actions of life, stimulating the consumption of cultural goods and knowledge for example. The uses of the sought and retrieved contents, whether empirical or abstract, are irrelevant. The construction of this informative value on searches will occur at the very moment when the search engine itself uses the content it will locate.

Note: Part of the conclusions of this article as well as additional details of the research presented here may be checked in GONZÁLEZ DE GÓMEZ, Maria Néida; RABELLO, Rodrigo (Org.). *Informação: agentes e intermediação*. Prefácio de Rafael Capurro. Brasília: IBICT, 2015. No prelo.

References

The references was made following the ABNT rules.

ALBAGLI, S. A questão do outro: do território ao Digital. In: GOMEZ, M. N. G. de; LIMA, C. R. M. [orgs.]. *Informação e democracia: a reflexão contemporânea da ética e da política*. Brasília: Instituto Brasileiro de Informação em Ciência e Tecnologia, 2010.

AUSTIN, J.L. *How to do Things witch words*. Oxford. Oxford University Press, 302p.

BARRETO, A. A. Recensão: Comunidade: a busca por segurança no mundo atual. *DataGramZero*, v.9 n.1 dez/08.

BARRETO, A. O mercado da informação no Brasil. *Revista Inf. & Inf.*, v.25, n.1, 2000.

BAUMAN, Z. *Modernidade líquida*. Rio de Janeiro: Zahar Ed., 2001.

BERNETT, D. VARVAKIS, G. Desafios das tecnologias de informação e comunicação sob a perspectiva da gestão do conhecimento na sociedade em redes. *DataGramZero - Revista de Ciência da Informação* - v.11 n.3 jun/10.

DEMO, P. Ambivalências da sociedade da informação. *Ciência da informação*, Brasília, v. 29, n. 2, 2000.

GRACIOSO, L. de S. *Filosofia da linguagem e Ciência da Informação: jogos de linguagem e ação comunicativa no contexto das ações de informação em tecnologias virtuais*. 2008. Tese (Doutorado em Ciência da Informação) - Instituto Brasileiro de Informação em Ciência e Tecnologia; Universidade Federal Fluminense, Rio de Janeiro. 2008.

LOPES, M. I. SILVA, E. L. A internet, a mediação e a desintermediação da informação. *DataGramaZero*, v. 12, n. 2, 2011.

MAIA, R. M. *Democracia e a internet como esfera pública virtual: aproximando as condições do discurso e da deliberação*. Disponível em: <http://vsites.unb.br/fac/comunicacaoepolitica/Rousiley2001.pdf>.

MARCONDES, D. A teoria dos atos de fala como concepção pragmática da linguagem. *Filosofia UNISINOS*, v.7, n. 3, p. 217-230, 2006.

MOURA; M. A.; GOMES, C. A. Comércio e consumo em tempo de conexões digitais: dimensões informacionais. *DataGramaZero - Revista de Informação* - v.15 n.2, 2014.
SALDANHA, G. S.; GRACIOSO, L. S. Filosofia da linguagem e Ciência da Informação na América Latina: apontamentos sobre pragmática e linguagem ordinária. In: RENDON, M. A. (coord.) *El problema Del language em la bibliotecología, ciência de la informacion, documentacion*. Un acercamiento filosófico-teórico. México: UNAM. Instituto de investigaciones bibliotecológicas e de La informacion, 1994.

ZILLER, J.; MOURA, M. A. Semiose e fluxos informacionais: os agenciamentos coletivos e a condição de usuário em ambientes digitais. *Liinc em Revista*, v.6, n.2, setembro, 2010.

Academic faculty formation in Archival Description



Laura Maria do Rego
São Paulo State University
lauramaria@marilia.unesp.br



José Augusto Chaves Guimarães
São Paulo State University
guima@marilia.unesp.br



Natália Bolfarini Tognoli
São Paulo State University
nataliatognoli@marilia.unesp.br

Introduction

The organization and representation activities of archival documents have their focus on the archival fund, that is, on the set of documents produced and/or received by the same individual or legal entity. The knowledge generated from the study of this fund is named archival knowledge.

One of the main activities of representing this knowledge is denominated archival description, whose function is to inform about the content and the context of the documents and their formal elements, with the purpose of giving access to the information contained in the funds, groups, series or documentary piece.

The description is, therefore, the highest point of the representation meeting the purpose of the documentation, which is to inform. The main task of archival make is a consequence of the organization of funds to turn documents available to society as

best as possible. Thus, the descriptive work aims to "make documentary funds accessible". (CRUZ MUNDET, 2001, p.255).

For Rodrigues (2003, p. 217), the description is a "[...] key function regarding the representation of archival information and the possibility of access to them", combining the observation of Tognoli (2012 , P. 82), in the sense that "[...] representation goes beyond the document, permeating its functions and its producing organs".

The description, inserted in the process of documentary management, has, as its object, the elaboration of representations of documentary or archival documentary groups, in which its function is established in the identification and representation of attributes that can "facilitate access, by users, to the contextualized information of its contents ", reports (HEREDIA HERRERA, 2011, p.8).

In this sense, Bellotto (2004, p.177) highlights the possibility of assessing a professional based on the work he/she develops in his/her documentary organization activities, since, for the author, a well elaborated elaboration of research instruments, will reflect in satisfactory information retrieval by users/ researchers, since poorly organized collection could ruin their work. Thus, the archivist, as a qualified professional, should act as an archival manager, developing, in his/her professional practice, description activities, considered nuclear in representation.

However, for this professional practice to be carried out in a systematic, consistent and methodologically defensible manner, a quality educational background is necessary for this professional to guarantee sufficient elements so that the archival documents can be treated in a way that they are accessible to the ones interested in their information, always respecting all the archival principles and providing these professionals with necessary knowledge about their ethical conduct as information professionals.

Concerns in Brazil regarding the creation of courses designed to enable professionals to become apt to treat archival documents began to emerge in 1911, with the approval of the National Archives (NA) through the Decree n. 9197, December 9, 1911. (MARQUES, 2008).

Archivists were professionalized through courses offered in library programs, and through the Permanent Course of Archives (PCA), offered in the NA, in order to meet the needs of the institution, but this reality begins to change with the

implementation of the First academic courses in Archivology in 1977. (MARQUES, 2007).

Since then several courses have been implemented in the country, in diverse regions, with the responsibility of providing the archivist with a multidisciplinary training, since we have professionals graduated in Archivology or coming from other related areas, such as Librarianship or History.

With the expansion of undergraduate courses in Archivology, notably in the last twenty years, Universities started to hire professionals with specific training and experience in Archivology, now with undergraduate, or with graduate courses in the area. However, the structure of Archivology courses does not always take into account the specificity of academic formation to carry out the attribution of disciplines among the teachers, which leads to situations of duplication or spraying of scientific production that lies between their formation in a specificity, and often does not establish a dialogue between them.

Thus, in order to discuss the articulation between the academic formation and the scientific production of the professors in Archivology courses in Brazil, specifically those related to the subject of description, from the triangulation developed by Guimarães (2003), teaching practice, academic formation and scientific production - although in the present research we will only address academic training and scientific production - the present study seeks to verify the formation and scientific production of the professors in Archivology courses, specifically in the disciplines that cope with the theme description, in order to verify to what extent the thematic specificity of faculty education is related to its pedagogical training and its scientific production.

Archival description professors in Archivology courses in Brazil: a study from lattes curricula.

There are sixteen undergraduate courses in archival science in Brazil. Thirteen courses are held at federal institutions and three at state institutions.

In this scenario, we were able to verify eighteen professors who teach subjects related to description. Out of these eighteen professors, eleven have a degree in Archival Science, six are graduated in History Degree/Bachelor, three in Librarianship, one in Business Administration and one in Law.

We verified the predominance of Archivology formation by the professors teaching disciplines related with archival description. This demonstrates that, over

almost 40 years, Brazilian Archivology courses are already responding significantly for the training of professors for the area. However, the formations in History and in Librarianship, act as important subsidies for the formation of these teachers.

It is also worth noting that among these professors, two have a dual formation in Law and Archivology, and Archivology and Full Degree in History, two professors have Bachelors degree in History and also a Full Degree, presented by Lattes as two formations. And there is still a graduate professor in Library Science who is studying Archivology.

As for graduate courses, at Masters' degree level, we observed that out of the eighteen identified professors, seventeen had this degree, as follows: nine had Master's in Librarianship/Information Science programs, three in History, two in Cultural Heritage, one in Education, one in Production Engineering and one professor with a Master's degree in progress in History.

It is possible to observe, therefore, the strong predominance of mastery in postgraduate programs in Information Science, which shows that graduate programs in Information Science, as well as ANCIB have been revealing, over the years, receptivity on archival issues.

At Doctoral level, which is fundamental for the professor, as evidenced by the full character of the researcher, it was observed that eight teachers had Doctorate degree, one in Knowledge dissemination, two in History, one in Education, one in Language, one in Social Memory, and two in Information Science. It is worth mentioning that out of these professors, six have a doctorate in progress, five of them are taking their Doctorate in graduate programs in Information Science, and one in Methodology and Research in Bibliotecas y Documentación, in a Spanish university.

With the characterization of the professors' formation in the disciplines related to the thematic archivist description in Archivology courses in Brazil, both undergraduate and graduate strictu sensu, we now analyze the scientific production of these professors, which constitutes in the set of documents generated from the results of research, published in different forms, and which represent the materialization of the knowledge generated on a subject of interest of a given scientific community. (DANUELLO, 2007, p.36-39),

Following the same context, the scientific production of these professors was analyzed, considering the following categories: book, book chapter, journal article and

complete papers published in annals of events, as these publications are the ones with greatest impact and visibility.

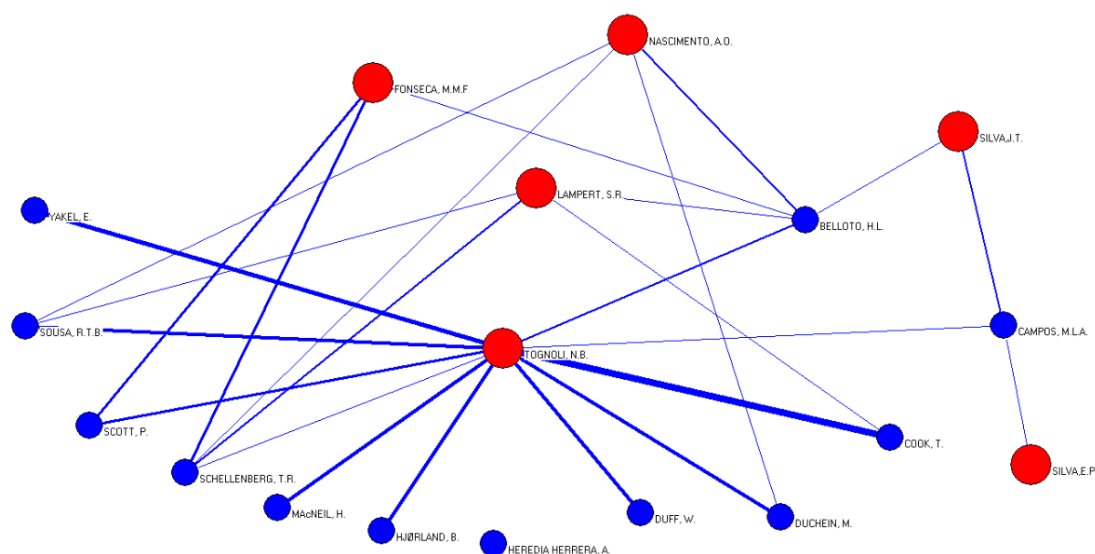
In order to collect these data, we identified all the productions available in these professors' lattes curricula, whether the titles of the publications presented the following terms: description; information representation; archival representation; knowledge representation, and for these terms we also identified the variation in English and Spanish. From there, we contacted the professors to request the materials to identify the cited authors in the productions. It is worth mentioning that some data were not presented, since the production was not current and the professors did not have them more in their collections.

From this collection, the references were analyzed, excluding papers with institutional authorship and auto-citations. Thus, we reached a set of 135 authors who received a total of 207 citations, which reveals an average of 1.6 citations per author.

For the purposes of analyzing these theoretical references, Price Elitism Law was applied, which states that, if k represents the total number of contributors in a discipline, \sqrt{k} would represent the elite of the studied area, as well as the number of contributors that generates half of all contributions (URBIZAGÁSTEGUI ALVARADO, 2008, p.1).

Thus, a total of 12 authors (9.6% of the total) who received 4 or more citations each, considering that this universe accounts for 29.8% of the total citations: Terry Cook, Heloísa Liberalli Bellotto, Theodore Roosevelt Schelenberg, Peter Scott, Renato Tarcisio Barbosa de Sousa, Michel Duchein, Elizabeth Yakel, Maria Luiza Campos, Wendy Duff, Antonia Heredia Herrera, Birger Hjolrand and Heather McNeil.

Based on this set of theoretical references, a citation network was built using PAJEK software, as follows:



Network 1: Most cited authors by professors in their productions.

Source: by the authors

The network allowed to observe the most cited authors by these professors in their productions, with the following data: Terry Cook (Canada) nine citations, Heloisa Liberalli Bellotto (Brazil) seven citations, Theodore Roosevelt Schellenberg (USA) seven citations, Peter Scott (Australia) six citations, Renato Tarcisio Barbosa de Sousa (Brazil) five citations, Michel Duchein (France) five citations, Elizabeth Yakel (USA) five citations, Maria Luísa de Almeida Campos (Brazil) five citations, Wendy Duff (Canada) four citations, Antonia Heredia Herreira (Spain) four citations, Birger Hjørland (Denmark) four citations, Heather MacNeil (Canada) four citations.

Conclusion

The present study showed a remarkable predominance of professor formation in Archivology courses, which demonstrates that the courses are managing to meet the demands for professor formation in the area. Regarding graduate courses, the predominance of courses in Information Science evidences, the receptivity of ANCIB to the archival issues, an approximation of Archivology and its professionals to the studies of Information Science, strengthening the relation between both to disciplines.

Concerning the scientific production of these professors, we verified that the five most cited authors reveal effective theoretical references insofar as they provide subsidies not only to teaching, but also to research in the area: Terry Cook, Heloísa Bellotto, Antonia Heredia Herrera, Renato Barbosa de Sousa and Michel Duchein, who provide a balance of theoretical influences, considering that authors from Canada, Spain, Brazil and France were identified, representing different theoretical currents.

References

The references was made following the ABNT rules.

BELLOTTO, H. L. *Arquivos Permanentes: tratamento documental*. 2a ed. Rio de Janeiro: FGV, 2004.

CRUZ MUNDET, J. R. *Manual de Archivística*. 4. ed. Madrid: Fundação Germán Sanchez Ruipéres, 2001.

DANUELLO, J. C. *Produção científica docente em tratamento temático da informação no Brasil: uma abordagem métrica como subsídio para a análise de domínio*. Marília: Unesp, 2007. Dissertação (Mestrado) – Programa de Pós-Graduação em Ciência da Informação - Universidade Estadual Paulista (Unesp).

GUIMARÃES, J. A. C.; DANUELLO, J. C.; MENEZES, P. J. Ensino de tratamento temático da informação (T.T.I.) nos cursos de Biblioteconomia do Mercosul: uma análise de capacitação e produção científica docente com vistas ao delineamento de políticas integradas para área. In: *ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO*, 5. 2003. Belo Horizonte. Anais. Belo Horizonte: Escola de Ciência da Informação da UFMG, 2003.

HEREDIA HERRERA, A. *Lenguaje y vocabulario archivísticos: algo más que un diccionario*. Sevilla, Junta de Andalucía, Consejería de Cultura, 2011.

MARQUES, A. A. C. *Os espaços e os diálogos da formação e configuração da Arquivística como disciplina no Brasil*. 2007. 298f. Dissertação (Mestrado) – Universidade de Brasília, Brasília, 2007.

MARQUES, A. A. C; RODRIGUES, G. M. Os cursos de arquivologia no Brasil: conquista de espaço acadêmico-institucional e delineamento de um campo científico. In: *XV Congresso Brasileiro de Arquivologia*, 2008, Goiânia. Congresso Brasileiro de Arquivologia, 2008.

RODRIGUES, G. M. A representação da informação em Arquivística: uma abordagem a partir da perspectiva da norma internacional de descrição arquivística. In: RODRIGUES; G.M.; LOPES, I.L. (Org.). *Organização e representação do conhecimento na perspectiva da ciência da informação*. Brasília: Thesaurus, 2003, v. 2, p. 210-230.

TOGNOLI, N.B. A representação na arquivística contemporânea. *Revista Ibero-Americana de Ciência da Informação*, vol. 5, n. 2, 2012.

URBIZAGÁSTEGUI ALVARADO, R. Elitismo na literatura sobre a produtividade dos autores. *Ciência da Informação*, v.38, n.2, Maio/Ago. 2009. Disponível em: <http://dx.doi.org/10.1590/S0100-19652009000200006>.

PROSPECTS IN KNOWLEDGE ORGANIZATION RESEARCH IN BRAZIL

Knowledge Organization and the core of Information Science: tales of big data, computing clouds and social networks

16



Renato Rocha Souza
Getulio Vargas Foundation
renato.souza@fgv.br



Mauricio Barcellos Almeida
Federal University of Minas Gerais
mba@eci.ufmg.br



Renata Abrantes Baracho Porto
Federal University of Minas Gerais
renatabaracho@eci.ufmg.br

Introduction

Within all scientific fields, and even among those more recently established, one can say that Information Science (IS) is one of the most introspective fields with

¹⁶ This work is partially supported by Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), Governo do Estado de Minas Gerais, Brazil, Rua Raul

Pompéia, nº101 - São Pedro, Belo Horizonte, MG, 30.330-080, Brazil

regards to research themes. The conceptual issues that underlie IS come up with very passionate arguments that their real research objects are sometimes blinded or faded into the background. Considering that seminal texts of the IS field are so broad – as seen in the works by Bush (1945) or Shannon and Weaver (1949) – and departing from such multifaceted concepts – again, as seen in Barlow (1994), Hofkirchner (1999) and Capurro (2003) – it is surprising that some authors are still able to find certain agreement about what IS should encompass under its umbrella (Borko, 1968; Zins, 2009). Indeed, the consolidation process of a relatively recent research field like IS, one finds many obstacles if compared to the same processes in the so-called “hard-sciences”.

With respect to this relatively unique characteristic in the IS field, Wersig (1993) emphasizes the need for a pragmatic-evolutionary perspective along with attempts at theoretical construction. This should be done so that both empirical research objects and interfaces with other fields, work as a support system for the permanent construction and re-construction of the whole research field. Wersig (1993) describes the IS professional through an analogy with the “weaving bird”. It is a kind of bird that builds its nest by weaving elaborate and complex nets, which refers to the thematic connections, required in the ordinary work of IS professionals.

As IS is a dynamic research field, working on research objects somehow fluid and ubiquitous, we present here an opinion about how the paradigmatic and epistemological transformation occurs. There is a need for this field when considering the scenario of new technologies and new ways to handle information. We openly assume our intention of producing a certain provocation, as long as we believe that there is an on-going deflation of IS field as an autonomous research field. This deflation has occurred because of the increasing overlapping among research objects of IS and other fields. Such phenomenon has resulted in a substantial reduction of interest in the scope of IS, as well as the migration to other fields, whose research objects could be considered legitimate IS research objects.

The remaining part of the article is organized according to three scenarios: the deflation, the transformation and future perspectives.

Deflation

Borko (1968, p.3), in its remarkable definition that attempts to capture IS epistemology and praxis, asserts:

Information Science is that discipline that investigates the properties and behavior of information, the forces governing the flow of information, and the means of processing information for optimum accessibility and usability. It is concerned with that body of knowledge relating to the origination, collection, organization, storage, retrieval, interpretation, transmission, transformation and utilization of information. This includes the investigation of information representations in both natural and artificial systems, the use of codes for efficient message transmission, and the study of information processing devices and techniques such as computers and their programming systems. It is an interdisciplinary science derived from and related to such fields as mathematics, logic, linguistics, psychology, computer technology, operations research, the graphic arts, communications, library science, management, and other similar fields.

Considering this definition, one might claim that virtually all research objects of the contemporary academic research would involve at least one component connected to IS. However, one can identify a bias, albeit broad, which can be read between the lines. The main concerns of IS are processes of information and knowledge representation, in addition to the manipulation of the records produced along these processes. The nature of these processes presupposes that information suffers successive abstractions, modeling and representations so that it can be organized, transmitted, codified and consumed. Buckland (1991) captures these processes and related objects by considering all of them as manifestations of the concept of information. Furthermore, it is worth emphasizing that information and knowledge representation are seminal subjects in the scope of IS.

Despite some discussion, there is a reasonable agreement about both the moment and the reasons for the birth of IS. On one hand, IS arises to deal with the issues that appeared from the increasing complexity in the management of informational collections. On the other hand, it fulfills the need for the creation and adaptation of methodologies – the praxis. Both aforementioned facts come from the fields of documentation and librarianship. However, the last few years of the 20th century were characterized by the fast expansion of information technology and by the acceleration of the processes of information transformation. These processes include the creation, representation, storage, organization, dissemination and consumption of

information. This meant the issues from decades ago originally justifying the appearance of a so-called “Information Science”, are nowadays exponentially felt.

Networks, mobile devices, tablets and other gadgets, as well as digital libraries and emergent developments like wearable computing, have continuously changed contexts, shorten cycles, and reinvented material supports. These events have redefined the relation between people and information records. There is a progressive “undocking” process, changing deeply the usual connections between information and its supports for recording, exchanging and consuming. Organization of the large mass of data needs new and creative solutions, suggesting the real need of a science to deal with information, namely IS.

This “technological ecology” could favor a new inspiration to the IS field, by multiplying the informational issues and the consequent demand for new solutions. But what has really occurred is the gradual migration of genuine IS research objects to other fields. To illustrate such thematic migration, one can mention socio-technical artifacts for information retrieval (like digital libraries), instruments for knowledge representation (like ontologies), techniques for domain modeling, information systems studies in their several contexts of use, among others. All of these examples of research objects have been actively studied by other sciences in fields of Linguistics, Administration, Computer Science and Information Systems.

On the other hand, themes like “social responsibility”, “inter and transdisciplinarity”, “information and work”, “information society”, and so forth (ANCIB, 2009), although relevant and legitimate themes for a social and applied science, have directing the field to focus on aspects related to the social issues coming from Sociology, Anthropology, History, among others. The rise of research in themes related to management, as Knowledge Management and Competitive Intelligence make IS close to fields of Administration and of some subfields of Engineering. All these examples show borderline overlapping that contribute to dissolving IS frontiers and to the lack of its identity.

With regard to the processes of information and knowledge representation, which are strongly connected to the origins of the IS field, one can expect that it become a mere and passive spectator in a highly dynamic environment. Even though IS is a dominant field when the subject is the study and construction of instruments for information organization, like indexing languages and information systems interfaces, these themes are nowadays associated with “technical sciences”. Indeed, these

sciences have incorporated these instruments into their constructs. Borko (1968) analyses the research objects mentioned in the “Current Research and Development in Scientific Documentation” and lists nine categories, namely:

- Information need and uses
- Document creation and copying
- Language analysis
- Translation
- Abstracting, classification, coding and indexing
- System design
- Analysis and evaluation
- Pattern recognition
- Adaptive systems

Among these objects, which are very current topics, only 1, 2 and part of 5 continue to be recognized as part of the core of IS. Linguistic analysis has been properly developed in fields like Computational Linguistics, Corpus Linguistics, Computer Science and even Applied Mathematics. Topics 3 and 4, as well as part of 5 and 7, have been developed under the label of Natural Language Processing. Computer Science and Engineering encompass topic 6, namely, Systems Design. Finally, the growing field of Artificial Intelligence has developed the topics 8 and 9.

In many universities around the world, the “schools of information science” have been transformed into “schools for information studies”, where special groups gather the required profiles able to exert the socio-technical interdisciplinary skills. In the United States, an example of this sort of discussion is the so-called *i-school’s movement*, in which several renowned universities have adopted new techniques, attitudes and even new curricula in schools and departments (Detlefsen, 2008). In addition, the conferences of the International Society for Knowledge Organization (ISKO 2014) also denote a clear direction that seeks a broad coverage, while maintaining the theoretical focus. The incorporation of themes that would suggest a social bias, as the *folksonomies*, occur in the scope of the knowledge organization systems, a very dear and traditional research subject within IS.

We conclude this first scenario emphasizing the need to take care of IS field, so that it does not become a mere niche among other fields. Some hope that information

professionals, instead of being the Wersig's weaving bird, become a sort of "remora"¹⁷, which feeds on the thematic leftovers of topics that other fields develop.

Transformation

There are surveys and research available in the literature, which characterize IS under countless different aspects, for example Capurro (1991) and Whittaker (2011), to mention a few. The scenarios presented here in this section do not correspond to an exhaustive literature review. Instead, we chose just to approach the effects of some new developments in IS, namely, Big Data, Cloud Computing and Social Networks. We believe that those developments may have caused impacts and paradigmatic changes in IS research.

Social Networks

Social networks are structures that gather actors – individuals and institutions – through links (Easley et al, 2010), which can be reified as social or technological arrangements. The representation and study of those networks include, but are not limited to, fields of Anthropology, Biology, Communication, Economy, Geography, Information Science, Psychology, Sociology and Socio-Linguistics.

Here, we only approach the social networks constructed from socio-technical arrangements, as is the case of some available on the web, for example, networks for relationships like Twitter¹⁸ and Facebook¹⁹. In its early days, Twitter was discarded because it was not considered a relevant information source, but in 2010 its economic messages of 140 characters started to be collected²⁰ by the United States Library of Congress. In 2014, more than 500 million messages were exchanged in Twitter each day²¹. Similarly, in 2014 Facebook had around 1.1 billion users exchanging an amazing volume of information²². Only these two instruments completely surpassed the

¹⁷ Parasite fishes that have on the head a sucking disk with which they attach themselves to sharks and other fishes so that they can take rests of food.

¹⁸ <https://twitter.com/>

¹⁹ <https://pt-br.facebook.com/>

²⁰ <http://www.businessinsider.com/library-of-congress-is-archiving-all-of-americas-tweets-2013-1>

²¹ <http://www.internetlivestats.com/twitter-statistics/>

²² <http://www.statisticbrain.com/facebook-statistics/>

information currently disseminated in printed or digital media. The main feature of this sort of information is the fluidity, dynamics and thematic ephemerality, which are explored in activities of environmental monitoring, sentiment analysis (Liu, 2012) and even the recording of human history.

It is worth mentioning another phenomenon, Wikipedia, one of the most remarkable crowd sourced collective developments of our age. Wikipedia is a free and collaborative multilingual encyclopedia, which makes it possible to copy, change and extend any type of information. It is distinguished from blogs because it does not use the traditional concept of authorship. Anyone with Internet access is able to modify any Wikipedia article, and each reader is a potential collaborator. In addition, Wikipedia goes beyond the results obtained by search engines, as the response to a query is a single text.

It is still worth remembering that social networks and their related phenomena point out to the ephemeral character of both the informational records and the instruments required to organize these records. Indeed, these collective constructions allow one to create products of unimaginable extent, which will require new approaches and studies (Surowiecki, 2004).

Big data

The explosion of information, which is sometimes called information or data deluge, is more than a large increase in the number of records. The term “big data” has become somewhat commonplace: it is used to refer to any mass of data, which cannot be processed without specific computational structures. In addition to the large volume, some definitions assigned to big data features like variety, velocity, and veracity (Laney, 2012). Indeed, since the 1990s large volumes of data have been manipulated with the aim of producing simulations and predictions. However, the textual non-structured information has been increasing a lot since the 2000s. This phenomenon has had a strong impact on the academic community and in research, most notably in fields like IS. The professionals using traditional techniques of subject analysis, classification and categorization have been trying to adapt their expertise, both to the gigantic collections, and to the speed of production of summaries and indexes (Magnusson; Vanharanta, 2003; Nodus Labs, 2012).

In the scope of academic communities, works of literature review, which are essential for research, are increasingly complex and in some cases impractical. Often a researcher has to cut arbitrary and explicit excerpts from the literature, meaning that the quantity of publications and available sources about specific subjects has become intractable. In other scientific cases, it is now a common research mode called “data driven research” (Moe-Behrens, 2012). According to this mode, one should perform several tests of hypotheses in large masses of data as an exploratory preamble. When promising possibilities are found from the data analysis, it would be possible to propose qualitative and quantitative enhancements though interactive changes of focus.

Cloud computing

Since we have already briefly described both the social phenomenon materialized through networks, and explosion of information illustrated through the hype around the big data, it is worth mentioning the importance of clouding computing. This phenomenon adds one more abstract level to the process of dematerialization of information records. Some time ago, still without confidence, people changed from physical supports to digital ones like tablets and mobile devices. But now, the cloud computing separates the users from those devices that store their data. Ubiquitous and often freely available storage services (DropBox²³, Google Drive²⁴, iCloud²⁵, to mention a few) are part of the lives of millions of people. These people seem unconcerned with the potential lack or theft of both photos and private data, even considering the current quantity of security incidents.

So, one can notice that: (a) social networks have disrupted the geographical limits of connectivity; (b) the new explosion of information has eliminated the possibility for an individualized treatment of information records; and (c) cloud computing has broken up the tangibility of records. So, in this complex context, an important issue begs for a proper answer: what will be the research agenda for IS in the 21st century?

Future Perspectives

²³ <https://www.dropbox.com/>

²⁴ <https://drive.google.com/>

²⁵ <https://www.icloud.com/>

What would be the role of IS in a context in which it seems to have lost primacy as a research field, and even centrality, in the thematic of information and knowledge representation? Resuming our initial provocation in the aforementioned scenario of IS deflation, what will be the possible arrangements capable of avoiding such deflation? Revisiting the definition of IS by Borko (1968), as “an interdisciplinary science that investigates the properties and behavior of information, the forces that govern the flow and use of information and the techniques of processing”, one can conclude that this program has become too broad for the IS field. Indeed, one can even conclude that such a program is too broad for any research field.

Yet, the already cited proposal of Wersig (1993) has become more and more present. It recognizes the new role of knowledge in contemporary society, and advocates the adoption of pragmatism to deal with problems in IS. Thus, in this scenario, epistemological approaches should receive less emphasis. Wersig’s third model for IS suggests, on one hand, a theoretical structure that reduces the attempts of formulating general laws. On the other hand, it suggests the creation of more action strategies from an approach based on the interlacement of scientific concepts. Then, it would weave a proto-net of basic concepts in IS, from which other people or groups could continuously find and integrate new concepts. In this sense, this conceptual net would become more inclusive and strong, in addition to having augmented it’s scientific character.

Examining this from another perspective, on one hand, interdisciplinary studies may weak IS as a research field. But, on the other hand, this same interdisciplinary studies may bring the possibility, and even the prerogative, of mediation among disciplinary dialogues. Such interdisciplinary essence fosters the professional and the researcher of IS to navigate in new theoretical spaces, to adapt to the technological contexts and to continuously reinvent itself.

Curriculum Vitae

Renato Rocha Souza holds a PhD in Information Science and currently acts as Adjunct Professor in the School of Applied Mathematics at the *Getulio Vargas* Foundation, Brazil.

Mauricio Barcellos Almeida holds a PhD in Information Science and currently acts as Associate Professor in the Department of Information Theory and Management at the Federal University of *Minas Gerais*, Brazil.

Renata Baracho Porto holds a PhD in Information Science and currently acts as Adjunct Professor in the Department of Information Theory and Management at the Federal University of *Minas Gerais*, Brazil.

References

ANCIB. (2009). History of the Brazilian Conferences of Research in Information Science. Retrieved July 31, 2011 from <http://www.ancib.org.br/enancib/historico-do-enancib/>

Borko, H. (1968). Information science: what is this? *American Documentation*, v.19, p.03-05.

Bush, V. (1945). As we may think, in the atlantic monthly. Retrieved August 22, 2010 from <http://ccat.sas.upenn.edu/jod/texts/vannevar.bush.html>

Capurro, R. (2003). The concept of information. *Annual Review of Information Science and Technology (ARIST)*, v.37, n.8, p.343-411.

Capurro, R. (1991). What is Information Science for? a philosophical reflection. In: Vakkari, P.; Cronin, B. (Ed.). *Conceptions of Library and Information Science. Proceedings...* London: Taylor Graham, p.82-96.

Detlefsen, E. G. What´s is a name? the I-Schools project. *Medical Library Association News*, v.28, n.3, p.1-17, 2008.

Easley, D.; Kleinberg, J. (2010). *Networks, Crowds, and Markets: reasoning about a highly connected world*. Cambridge: University Press.

Hofkirchner, W. (1999). The quest for a unified theory of information. In: Hofkirchner, W. (Ed.). *Proceedings of the 2nd Intern. Conf. on the Foundations of Information Science*. Amsterdam: Gordon & Breach, p. 9-30.

Ingwersen, P.; Järvelin, K.. (2005). *The turn: integration of information seeking and retrieval in context*. Netherlands: Springer, v. 18.

International Society For Knowledge Organization. ISKO Chapters. Retrieved March 29, 2014 from <http://www.isko.org/chapters.html>

Laney, D. (2012). The Importance of 'Big Data': a definition. Retrieved March 22, 2014 from <https://www.gartner.com/doc/2057415/importance-big-data-definition>

Liu, B. (2012). *Sentiment Analysis and Opinion Mining*, Chicago: Morgan & Claypool.

Magnusson, C.; Vanharanta, H. (2003). Visualizing sequences of texts using collocational networks. In: *3rd Intern. Conf. Leipzig*. v.2734, p. 276-283.

Moe-Behrens, G. (2012). From hypothesis to data driven research. Retrieved April 2, 2014 from <http://www.sciencetogrok.com/2012/12/from-hypothesis-to-data-driven-research.html>

Nodus Labs. (2012). Texttexture: visualize any text as a network. Retrieved July 11, 2013 from <http://texttexture.com/>

Osterberg, G. Update on the Twitter archive at the Library of Congress. Retrieved July 3, 2014 from <http://blogs.loc.gov/loc/2013/01/update-on-the-twitter-archive-at-the-library-of-congress>

Shannon, C.E.; Weaver, W. (1949). *The Mathematical theory of communication*. Urbana: University of Illinois Press.

Surowiecki, J. (2004). *The Wisdom of Crowds: why the many are smarter than the few and how collective wisdom shapes business, economies, societies and nations*. New York: Anchor.

Vakkari, P.; Cronin, B. (1992). *Conceptions of Library and Information Science. Historical, empirical and theoretical perspectives*. London: Taylor Graham.

Wersig, G. (1993). Information science: the study of postmodern knowledge usage. *Information Processing & Management*, v.29, n.2, p.229-239.

Whittaker, S. (2011). Personal information management: from information consumption to curation. *Annual Review of Information Science and Technology*. v.45, n.1, p. 1-62, 2011.

Zins, C. (2009). Knowledge map of information science: issues, principles, implications. Retrieved September 30, 2012 from <http://www.success.co.il/is/index.html>

Knowledge Organization: research and development



Gercina A. B. O. LIMA
Federal University of Minas Gerais
limagercina@gmail.com

1 Introduction

This paper aims to examine Knowledge Organization Research trends in Brazil. First of all, it's necessary to clarify some characteristics and overlaps according to areas related to the field of study and its approach, such as Classification, Knowledge Organization (KO), Library and Information Science (LIS).

Knowledge organization has its historic roots based on theoretical basis of Classification. Classification is a subject field that classifies philosophical knowledge that begins with the Categorical Systems of Aristotle as well as a field of Library Classification which has processes and tools used to represent thematic content of documents, as well as increasing effectiveness of information retrieval. In fact, the two approaches are intended to relate theory and practice in studies and researches in the field of Library and Information Science (LIS). Knowledge Organization (KO) is also considered a field of study and an activity as well as Classification. Both KO and Classification are some of the main subject fields of LIS.

According to Dalhberg (2006), Knowledge Organization as a field of study is related to the creation of International Society of Knowledge Organization (ISKO), on 22 July 1989.

Dalhberg (1993) clarifies that the term Knowledge Organization had been already used by Henry Evelyn Bliss to create titles of his Works, such as *The Organization of Knowledge Organization and the system of the science*, *The Organization of Knowledge Organization in Libraries*, published respectively in 1926 and 1936. This term was also used in Dagobert Soergel's title of the thesis entitled *Organization of knowledge and documentation* (1971), defended in Germany as well as in the author's

thesis, which was published and entitled *Foundation of Universal Organization of Knowledge*.

The international journal *Knowledge Organization* (KO), published under the responsibility of ISKO, was published for the first time in 1993, replacing *International Classification*, which was published from 1974 to 1992. From there on, Dalhberg (1993) explains that the members of ISKO decided that the term Classification would be interpreted as a classification method, which groups together similar objects. However, The Knowledge Organization Journal aims to spread new ideas of Works that contributes to the development of new knowledge about Concept theory, Classification, Indexing and Knowledge Representation. It can be noted the close relationship between Classification and Knowledge Representation, both in theoretical basis and practice.

McIlwaine and Williamson (1992) claim that some researched areas are not as new as it is supposed. According to them the theory and the principles of classification will be always a fundamental point in all aspects related to Knowledge Organization.

Hagar Gomes (2009), in his paper *Trends in Knowledge Organization Research* points out that:

It can be noticed that KO is a subject field based on multiple ways to practice. Researches in this area, specifically, show us, at least, two major aspects: on the one hand, theoretical and philosophical approach focused on epistemological questions in order to analyze a domain together with its applications in the organization of classification systems and schemes, on the other hand, methods employed in the compilation of systems, which has led a number of authors to consider them as additional activities, despite the strong theoretical basis.

The author also adds that Knowledge Organization (KO) and Library and Information Science (LIS) have a close relationship, which led us to consider that the existing literature is confused between theoretical basis of KO and LIS. In other words, it's not sufficiently well defined if activities belong to KO subject field or LIS one, or if they are related to one another. Perhaps that is why Smiraglia (2005, 2006), editor of Knowledge Organization Journal, demonstrates his concern when he states that "[...] *what precisaly KO is [...] although many have written about it.*", (2005. p.139), which shows that there is no single definition in this area, going back to the original point when KO was called Classification.

Another relevant observation, concerning the use of the words Knowledge Organization (KO) and Information Organization (IO) have been used in different contexts, by several authors in the field. Bräscher and Café (2010) present a study on these words in the field of Information Science (IS). According to the authors the concepts are not properly delimited. The authors state that:

Knowledge Organization is sometimes a term used in the sense of Information Organization and vice versa. In certain situations, these terms are used together, information and knowledge organization. (BRACHER; CAFÉ, 2010, p.85).

According to the same authors, there are two types of organization processes: Information Organization, designed to describe the physical and conceptual characteristics of information objects, near activities developed in libraries and documentation centers; and the Knowledge Organization, designed to the creation of models that can represent knowledge units of a particular domain, according to its characteristics. In this way, they point out that

These two processes (IO and KO) produce two different types of representation, information representation understood as a set of attributes that represents a particular information object, obtained by the process of physical and content description, and Knowledge representation, built into a conceptual framework that represents models of World.(BRACHER; CAFÉ, 2010, p.93).

It can be said that, these two processes complement each other. On the one hand, if there is a content represent process of a specific document in the IO field, which first of all has a recorded knowledge and, subsequently, must be retrieved by user in an information retrieval (IR) System, On the other hand, KO field organizes knowledge by using a domain models, representing its structure based on description of concepts and semantic relationship among them. The structures are usually created with the help of scientific experts, whose knowledge are validated and organized based on their expertises.

This present research has considered works whose trends are based on Knowledge Organization Field. However, they were not differentiated from these other fields because there are no standard for most of titles of published works, which has already reported in the literature.

First of all, in order to map these trends, it has been proposed to do a brief review of research in the OK field by searching journal articles and proceedings of national and international events, without trying to exhaust the search. It is emphasized, therefore, that the result of the bibliographic research was conducted between 1993 and 2015.

The bibliographical research was done by using the following expressions, in English and in Portuguese: Trends in Knowledge Organization Research, Information Organization Research, Information Organization and Representation in Brazil, teaching and research in Information Science, current trends in Knowledge Organization, Knowledge Organization research, trends in classification research.

2 Trends in Knowledge Organization Research through time: a summarization

Any research in a particular field of knowledge advances from questions, which consider historical basis and intellectual experience already developed by researchers in the field. These approaches increase our understanding of researches that have already been developed and also help us to understand current activities, allowing the creation of knowledge gaps which need to be better studied and relevant questions to be developed.

First of all, for presentation of selected works whose trends are in Knowledge organization, it has opted to describe in a summarized way, firstly, articles of foreign researchers and based on the field, secondly, articles of Brazilian researchers, according to the chronological order of presentation, and finally to expose relevant themes based on researches that may contribute significantly to this field in Brazil currently.

Already in 1981, Elaine Svenonius expressed a concern about research in the field, at that time called bibliographic control. In her article called "Directions for research in indexing, classification, and cataloging", the author points out the need to research in the field to solve problems of Knowledge/Information Organization in order to improve its retrieval. The author's article suggests the following studies: (1) the difficulty of measuring the performance of information retrieval in information retrieval (IR) Systems, because it is strongly dependent on its relevance and recall and precision measures have not been helped because they do not always consider user

needs, (2) need to improve semantic relations for improvement of indexing languages, (3) need to improve automatic indexing by developing the most efficient extraction techniques, – in this case, it is suggested the contribution of Computational Linguistics and uses of artificial intelligence; (4) need for more theoretical studies to assess the library classification system in order to improve the adaptability and the hospitality, mainly in specific domains, (5) researches of compatibility, conversion and translation of terms, which are controlled to solve problems of interoperability of systems.

The paper “Current trends in knowledge organization” was published by Ingetraut Dahlberg, in 1993. In this paper, the author presents state of the art about the advances and trends in Knowledge Organization research. After the historical introduction to the area, Dahlberg presents a systematic approach based on literature in the area in order to verify what research trends would be studied in that moment. The findings pointed out the need for research based on: (1) automation of Knowledge Organization, including automation of universal systems of classification; (2) studies of creation of universal thesaurus, based on linguistic dictionaries or based on the faceted structure; (3) compatibility between library classification system and the existing thesauri, mainly in order to meet the demands in different domains of knowledge; (4) development of structures for library classification system and thesauri by using studies of categories and their relations; (5) study on using object oriented programming; (6) improvement of quality of indexing process and subject cataloging; (7) needs for promoting links between professionals of knowledge organization and terminologists, in order to use the methodologies of terminology in the definition of the concepts.

Hjørland and Albrechtsen (1999) published a study called “An analysis of some trends in classification research”, in which they analyze three aspects presented by the then president of ISKO, Clare Begthol (1998). According to the author, these three important points are fundamental in Knowledge Organization field:

1) the academic disciplines as the main structural principle; 2) the fiction/nonfiction distinction as one secondary structural principle; and 3) information retrieval techniques that call into a question whether a whole document (e.g; book, article) is the most appropriate unity of analysis in online retrieval systems. (BEGTHOL, 1998, p.1).

The authors consider these three points based on an epistemological analysis and points out that: (1) The academic disciplines provide principles for central or

fundamental structure in division system of classification schemes; (2) appreciate how knowledge classification can represent some specific fields of knowledge, without losing guaranteeing cultural and epistemological aspect, therefore the library classification systems must reflect on interdisciplinary fields and relation between fields of knowledge; (3) Regarding division of knowledge in fiction and nonfiction, Hjørland and Albrechtsen (1999) recognize that this point is not a fundamental question, however there are some types of knowledge, mainly in the area of humanities, such as Psychology, for example, that must be represented in different aspects; (4) indexing of chapters and articles is already a common practice in information retrieval system, therefore the users need to access documents, the whole or parts of the document, by using chapters or journals, and it is important to consider that units of analysis are related to content, in other words they are not related to structure of document; (5) need for studies that improve the information access points to bibliographical databases or full texts in order to optimize information retrieval.

Also in 1999, McIlwaine and Williamson present the paper called “International trends in subject analysis research”, in which describe the research conducted on the research theme subject analysis, for a period of 10 years: from 1988 to 1998. The authors’ research has focused on three areas: (1) universal library classification system, (2) thesauri development and design, and (3) efforts to organize internet. The authors point out that: (1) the progress of conversion of CDD, CDU and LCC system, widely used, in Machine-readable format constitute a great advance in universal system of classification; (2) these systems can be used to improve information retrieval on internet. With respect to thesauri, the main research topic related was the use of a faceted analysis for thesauri construction; (3) Thesaurus has started to be studied as a tool for organizing and navigating to be used on internet; (4) need for creating a multilingual thesaurus, not with the intention of translating terms, but with the intention of conducting surveys that allow compatibility between thesauri and identification of methodologies, making possible the access and manipulation on line; (5) classification process has already been used in some websites, though made in a less developed manner. However, it is required studies based on interactive interfaces, citation order, data mining, thesauri in hypertext format and intelligent agents.

In 2003, McIlwaine presents the paper called “Trends in Knowledge organization research” and points out that: (1) interoperability of systems is one of the most important topics to improve the access to the internet through the artificial







intelligence approach by using linguistic, such as natural language processing; (2) Most of surveys are conducted at University, by using government resources, available for a short period of time, which makes survey development and its conclusion difficult; (3) Regarding the problem of language, the author suggests the creation of vocabularies linked to a verbal system, preferably a multilingual one, carried out with the assistance of a number of linguistic and artificial intelligence experts, in order to solve the problem of interoperability, multilingualism and interactive interface for user.

In 2008, Lopèz-Heurtas presents the paper called “Some currents research questions in the field of knowledge organization” an extensive review of the literature and points out some research needs: (1) improve the quality of knowledge representation and organization and information retrieval on internet environment; (2) the multilingualism for knowledge organization; (3) studies based on integrating sociocultural differences for knowledge organization in order to improve the quality of knowledge organization systems (KOS); (4) better representation and integration between sectors recognized as marginalized and minority sectors in knowledge organization systems; (5) integration of knowledge to construction of KOS in order to ease the integration of several structures on web; (6) multidimensional knowledge organization and representation in KOS; (7) use of domain analysis for knowledge and management in interdisciplinary and transdisciplinary places, using bibliometric approach and domain analysis terminology; (8) creation of ontologies and interdisciplinary systems.

The paper “Knowledge organization trends in library and information studies: a preliminary comparison of the pre- and post-web eras”, by Saumure and Shiri published in 2008, makes reference from 1966 to 2006. The qualitative analysis done could identify the following themes which represents the research trend in pre-and post-web eras, based on an analysis of 219 publications: (1) organizing corporate or business information; (2) machine assisted knowledge organization, (3) librarians as knowledge organizers of the web, (4) interoperability, (5) cataloging and classification, (6) classifying Web information, (7) Digital preservation and libraries, (8) metadata applications and uses, (9) cognitive models, (10) education, (11) indexing and abstracting, (12) Thesauri initiatives.

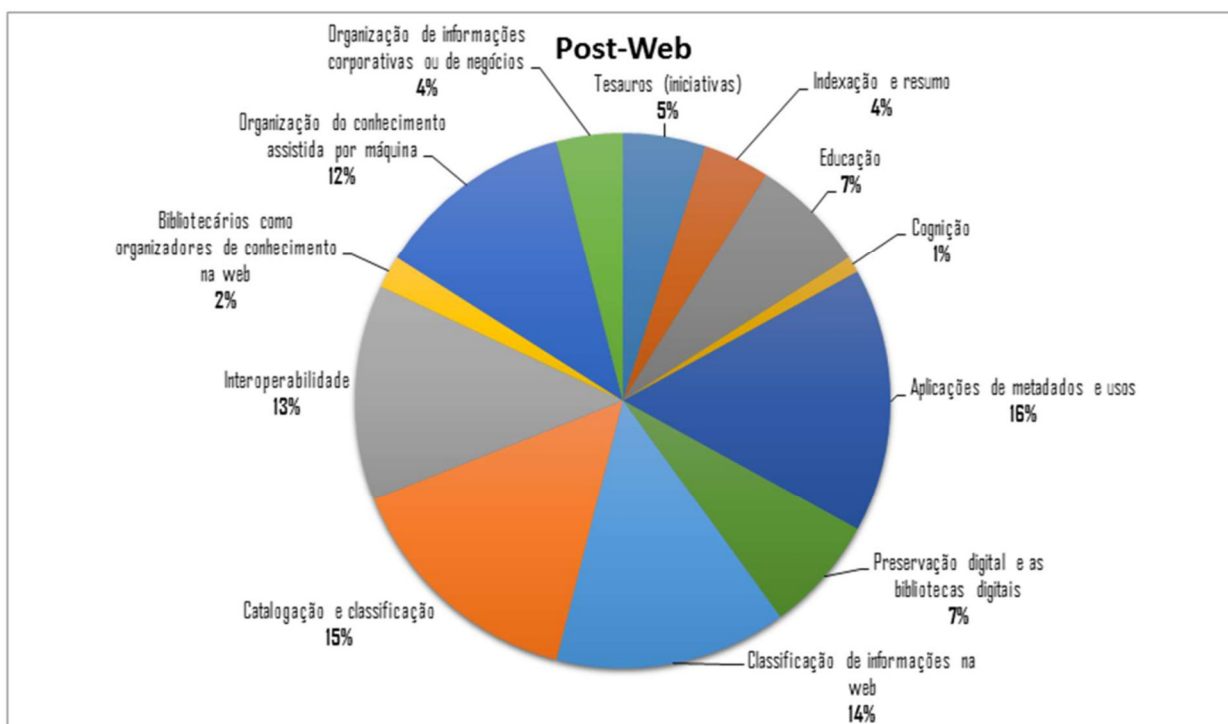
The trends in the 31 selected articles representing pre-web era (GRAPHIC 1) and the 188 selected articles representing post-web era (GRAPHIC 2), are presented in the graphic below:



	29% Indexing and abstracting
	7% Thesauri initiatives
	14% Machine assisted knowledge organization
	14% Cataloguing and classification
	29% Cognitive models
	7% Education

Graphic 1 – Numerical Representation of pre-web trends

Fonte: translated by Saumure and Shiri (2008, p.654).



Graphic 2

Source: translated by Saumure and Shiri (2008, p.654).

- (5%) Thesauri initiatives.
- (4%) Indexing and abstracting,
- (7%) Education,
- (1%) Cognitive models,
- (16%) Metadata applications and uses,
- (7%) Digital preservation and libraries,
- (14%) Classifying Web information,
- (15%) Cataloging and classification,
- (13%) Interoperability,
- (2%) Librarians as knowledge organizers of the web,
- (12%) Machine assisted knowledge organization,
- (4%) Organizing corporate or business information;

In 2010, Ibekwe-sanjuan and Sanjuan published the paper “Knowledge organization research in the last two decades: 1988-2008”, in which they reviewed the literature, by citing researches from present study: Hjørland and Albrechsten (1999), Mcilwaine (2003), Lopez-huertas (2008) and Saumure and Shiri (2008). The authors report that most of these studies did not make clear the methodology used, in this way,

reading and interpretation of the data were made on the basis of knowledge of mentioned authors above, who are responsible for reading and analysis besides inferring and arriving conclusions. This survey was conducted by using bibliometric methods and data analysis in order to identify trends in knowledge organization in 31 international journals on Web of Science (WoS), separated by two decades. As a result of the first decade (1988-1997), the authors achieved the following trends: (1) knowledge, (2) classification, (3) knowledge organization, (4) vocabulary control, (5) design of bibliographic databases, (6) indexing, (7) thesaurus construction and usage, (8) information and text analysis, (9) information-documentation, (10) Information Science, (11) knowledge representation and organization (11) classification schemes. In the second decade (1998-2008), the following trends were mapped: (1) classification, (2) information, (3) knowledge, (4) knowledge organization based on its different aspects (transfer, perspective, domain, integration, management, flow, map, recall, representation, network and literature), (5) Information science, (6) epistemological foundations, (7) terminology database, (8) information retrieval, (9) metadata, (10) gay-lesbian classification vocabulary, specifically developed for describing publications on homosexuality, (11) Web. It is important to remind that the last three topics were not present in the first decade.

The last international paper to be presented is published by Beak et al (2013), called “International comparative domain analysis in knowledge organization research topics in four countries - Brazil, South Korea, Spain and the United States”. Each author has chosen two journals of Library and Information Science (LIS) in the country of each one of them. The authors present statistical analysis, comparing the terms overlap each other in the four countries, the terms that appear only in one of the countries and also analyze their occurrence in each country relation to other countries. The occurrence of the terms in each country is shown by the table 1 bellow (it is extracted from the paper).

Graphic 2 – Top frequent keywords from each country

Source: translated by author of BEAK, J. et al, 2013. p.34.

Brazil	South Korea	Spain	United States
---------------	--------------------	--------------	----------------------

Knowledge representation	Korean Decimal Classification	Semantic Web	Classification
Knowledge management	Dewey Decimal Classification	ontologies	Information retrieval
ontologies	RDA	Thesauri	Information Science
Automatic indexing	<i>Nippon</i> Decimal Classification	Knowledge organization	Systems
Classification systems	Cataloging rules	Information architecture	Retrieval
Information organization	FRBR	Knowledge management	Science
Information science	OPAC	Information retrieval	<i>Web</i>
Knowledge Management tools	Library catalogs	Information Systems	Models
Online catalogs	Metadata	internet	Information
Semiotics	Subject headings	Knowledge organization systems	Knowledge organization
Terminologies		Legal Information systems	Search
		Metadata	Categorization
		SKOS	Knowledge
		Spain	
		University libraries	
		<i>Web 2.0</i>	

One of the first surveys conducted in Brazil on Knowledge organization was published in 2008, by Mariângela Fujita, under the title of “knowledge organization and representation in Brazil: analysis and conceptual aspects and scientific production of ENANCIB in the period from 2005 to 2007”. In this paper, Fujita (2008) presents an overview of research in Brazil, identifying themes and lines of researches, theoretical

and methodological foundations and scientific development in the field. To that end, it is conducted a systematic study on scientific production of Knowledge Organization and Representation (KOR) Research Group (RG2/GT2), of National Meeting of National Association for Research (ENANCIB), held annually by National Association for research and postgraduate in Information Science (ANCIB), conducted from 2005 to 2007. The author's analysis points out the incipience of scientific studies connected with interdisciplinary fields and also, stresses the emergence of themes such as ontology, taxonomy, semantic web and need for further studies on classification system in order to make them more flexible to meet the requirements in this area.

Also in 2008, Guimarães presented the paper called "The theoretical dimension of subject information treatment and its possible dialogues with the scientific universe of ISKO", in which analyzed knowledge organization dialogue with ISKO, from three theoretical approaches: subject cataloging, under North American influence; indexing, under English influence; and documentary analysis, under French influence. Regarding this context, the author analyzes researches and studies presented in RG2(GT2) - ANCIB, and also the ISKO Brazilian Chapter and he concludes that studies conducted at Brazilian schools of LIS, in this field of study, have pointed towards equity for studies and researches realization, without losing sight of international approaches. In view of the need to develop the field, Guimarães (2008) points out: (1) need for reflection on theoretical and methodological aspects, covering interdisciplinary dimension without losing the main focus on processing information in order to achieve an efficient retrieval; (2) Reflect on professional activity in face of this need, and (3) suggests that in context of ISKO, it must not be lost of sight how subject information treatment is making its theoretical and methodological framework.

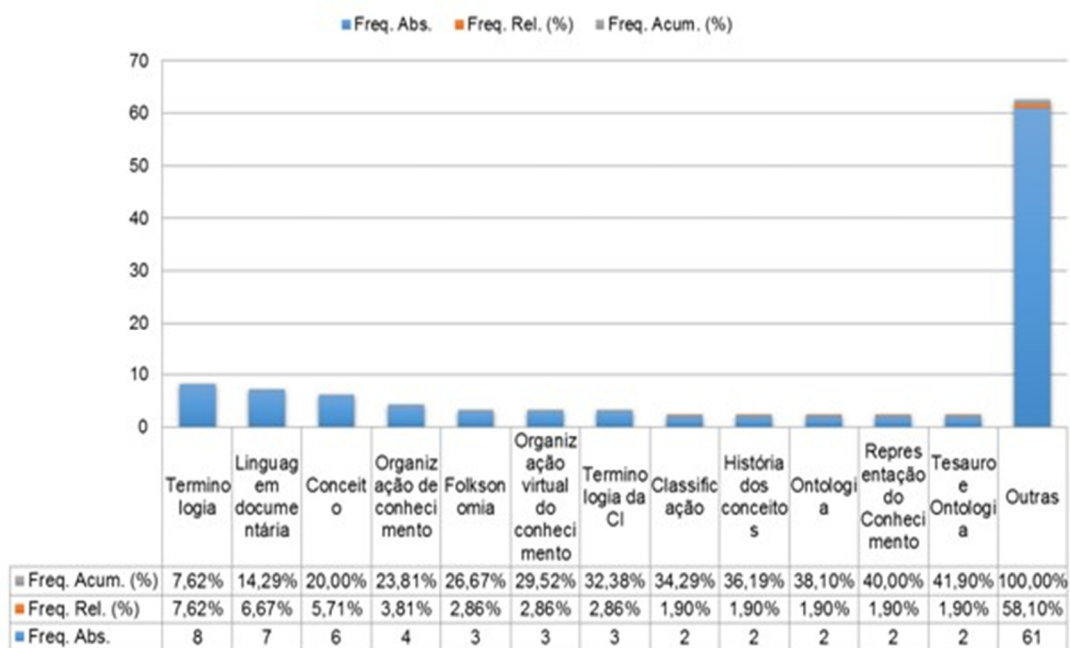
Hagar Espanha Gomes published in 2009 the paper called " Trends of research in Knowledge organization", based on a lecture held in the II Research Seminar in Information Science, organized by Information Technology, Search and Education Coordination of IBICT, in Rio de Janeiro, in 2007. The author describes a general overview in the field and points out two important aspects in the research: (1) theoretical and philosophical approach focused on epistemological questions in order to analyze a domain that aims to develop systems and classification schemes, and (2) methods used to build these systems. In addition, Gomes (2009) highlights other important points that should be considered: (1) needs for developing a theoretical sound basis for activities in LIS to enable a more comprehensive theory that can

subsidise surveys between theory and practice; (2) the multidisciplinary and interrelationship that occur between fields of research, increasing the need to study traditional classification schemes and the development of other systems, making them as hospitable/flexible as possible, in order to allow the creation of new concepts in this field, (3) suggests the domain analysis approach, the theory of integrative levels, and the faceted methods used as an alternative to these processes, (4) researches on the use of Dewey Decimal Classification (DDC) and the Universal Decimal Classification (UDC) as a tool for supporting web information retrieval and representation.

In 2014, Guimarães et al. Published the paper called “ The subject dimension analysis of knowledge organization research: a domain analysis of the Brazilian, Spanish and North American ISKO chapters’ conference”, in which the authors conduct an exploratory research based on domain analysis of articles published in ISKO chapters’ conference - Brazil, ISKO - Spain-Portugal and ISKO - Canada-United States in 2011 and 2013. Based on analysis of articles, the authors pointed out the predominant subjects in the three chapters: (1) epistemology, (2) domain analysis, (3) indexing language and process, including classification.

The last article found in Brazilian literature, in bibliographical research conducted here was “Authorship and research on knowledge organization: an analysis of the scientific production on Information Science”, published by Bufrem; Freitas; Nascimento (2014). The authors conducted a citation and co-citation analysis, using a corpus of 105 articles in the field, published from 2003 to 2012. These surveys present tables that are used to show which are the most fruitful journals in the field of knowledge organization, the most representative authors cited in the scientific production in the field, the approach of the research found in the corpus of the articles and the most common topics in this corpus. This last item will be presented in the table below (TABLE 1).

Table 1 – The most common topics in the corpus of knowledge organization articles in Brazil (2003 – 2012)



Source: (BUFREM; FREITAS; NASCIMENTO, 2014, p.160), adapted by the author.

Terminology
Documentary language
Concept
Knowledge organization
Folksonomy
Virtual Organization of knowledge
LIS terminology
Classification
History of concepts
Ontology
Knowledge representation
Ontology and thesaurus
Other

According to this brief literature review above, it can be stated that theory and research on Knowledge organization are cyclical processes because theory needs search for its development and application, at the same time search needs theory for

supporting the study, defining and improving future works which could lead to significant steps forward in the field. Svenonius (1981) uses the following thought to conclude her paper: "...theory without practice is empty, but practice without theory is blind".

2.1 Current trends on knowledge organization research in Brazil

Computers have brought a lot of benefits for knowledge organization field, contributing to make storage, processing and retrieval information easier. Knowledge Organization researches have had contributions to development in the field, mainly because of intersection of fields such as Computer Science, Linguistic, Terminology and Cognitive Science, which dates back to the 1950s. This progress had an impact firstly on activities related to indexing process with the advent of automated systems to create indexes and automatic processing of text, with linguistic contribution, in cataloging process, with the creation of MARC format, by Library of Congress (LC), which is effectively used in 1966, and considered by International Standard Organization (ISO) to be the first international automation project that establishes standards for bibliographical description in a machine-readable form, thus enabling data exchange. Therefore, classification process can presume the possibility of creation and compatibility of controlled vocabularies with the help of computer. All these procedures directly impacted on search and retrieval information within the information retrieval (IR) systems.

Since then, several studies and experiments have been performed, as exemplified above. It should be noted that changes in the traditional way of organizing knowledge had an ever stronger impact from the consolidation of the internet in 1980s.

However, this work does not aim to present the entire history of these events. In addition, not all previous research questions have been developed and also have not found satisfactory responses. It is therefore essential that these questions should become a pertinent point for research. On the basis of the foregoing, it should draw attention to current needs in this field of study, mainly in Brazil:

- 1) Theory and methods of Knowledge Organization need a better articulation and basis to achieve more consistency and coherence in this field of study;

2) The interdisciplinarity in the field requires a terminology mapping in order to create an epistemological field, avoiding ambiguity, since a variety of terms are used in intersection fields;

3) Searches on possible solutions for interoperability, aiming the sharing of information resources with especial attention to development of metadata standards and schemes in order to make easier information access to several digital repositories published on the Web;

4) Searches on the potential for application of faceted classification in organization and representation, as well as in web retrieval information;

5) Searches on use of Bibliographical classification systems (DDC and UDC) used as potential tools of navigation and search, by using of its hierarchical structures as alternative of information retrieval;

6) Researches which enable the compatibility between different thesauri, as well as the identification of methods which enable they can be used, accessed and manipulated online;

7) Researches on how to model and represent the multidisciplinary of knowledge domains in Knowledge Organization Systems (KOS) considering user needs;

8) Further researches on contribution from Terminology to construction of KOS in order to mold the specific domain before its representation, with the aid of conceptual definitions used to establish correlations between concepts from different fields of knowledge;

9) Searches on application in the information retrieval in different contexts in order to support the exploratory research;

10) Research on the potential of data-model "Simple Knowledge Organization System (SKOS)" used to represent the structure of different types of KOS, enabling this tool to be published with its data interlinked by Linked Open Data;

11) Greater integration between communities of knowledge Organization professional and fields of intersection;

12) Evaluation of search of keywords versus search with help of controlled vocabulary in a web environment;

13) Researches on implementation of RDA (Resource Description and Access), which allows the representation and description of digital resources.

Researches on automatic indexing, developing more efficient techniques for term extraction with contribution of Computational Linguistics and use of artificial intelligence.

3 Final Considerations

It can be stated that technological advances from 1950 until the present day have been relevant to development of Knowledge Organization field both in information storage and representation and information retrieval. The storage, previously made only in computer hard disk with large processing power, nowadays can be made in different ways, with a possibility of accessing files, data and applications from anywhere, at anytime via a computer or mobile device, since there is internet connection with cloud storage. Recognizing that development and growth of internet are ever-faster, new research interests arise in Semantic Web and Web 2.0 technologies, with a greater degree of involvement of users, which contributes not only to provide metadata but also to social bookmarks such as folksonomy. User needs would be the central point of this field of study, since information retrieval is the main focus of all field of knowledge Organization. We must not just consider the dependency on technological development, but also the greater integration between Knowledge Organization professionals and professionals of interdisciplinary fields. Accordingly, the greater the degree of integration of these professionals, the greater the degree of advances in researches.

References

The references was made following the ABNT rules.

BEAK, J. et al. International comparative domains analysis in knowledge organization research topics in four countries - Brazil, South Korea, Spain and the United States. *North American Symposium of Knowledge Organization*, v. 4, n.1, p.30-39. 2013. Disponível em: <<https://journals.lib.washington.edu/index.php/nasko/issue/view/1012>>. Acesso em: 2 ago. 2015.

BEGHTOL, C. Knowledge domain: multidisciplinary and bibliographic classification systems. *Knowledge Organization*, v.25, n.1/2, p.1-12. 1998.

BRÄSCHER, M.; CAFÉ, L. Organização da Informação ou Organização do Conhecimento? In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 9, 2008, São Paulo, Anais. São Paulo: ANCIB, 2008. Disponível em: <<http://www.enancib2008.com.br>>. Acesso em: 31 jul. 2015.

BUFREM, L. S.; FREITAS J. L.; NASCIMENTO, B. S. Autoria e pesquisa em Organização do Conhecimento: análise da produção científica em Ciência da Informação. *Revista da Faculdade de Biblioteconomia e Comunicação da UFRGS*, v.20, n.3, p.150-165. Ed. Especial, 2014.

DAHLBERG, I. Current trends in knowledge organization. In: MARCO, F. J. García (Coord.). *Organización del conocimiento en sistemas de información y documentación 1*. Actas del I Encuentro de ISKO-España, p.7-25, 1993. Madrid: Librería General.

DAHLBERG, I. Knowledge organization: a new science? *Knowledge Organization*, n.33, v.1, p.11-19. 2006.

DAHLBERG, I. Knowledge organization: its scopes and possibilities. *Knowledge Organization*, Frankfurt, v.20, n. 4, p.211-222, 1993.

DIAS, E. W. Ensino e pesquisa em ciência da informação. *DataGramaZero - Revista de Ciência da Informação*, v.3, n.5, out./ 2002. Disponível em: <http://www.dgz.org.br/out02/F_I_art.htm>. Acesso em: 2 ago. 2015.

FUJITA, M. L. S. A. Organização e representação do conhecimento no Brasil: análise de aspectos conceituais e da produção científica do ENANCIB no período de 2005 a 2007. *Tendências da Pesquisa Brasileira em Ciência da Informação*, v. 1, n.1. 2008. Disponível em: <<http://inseer.ibict.br/ancib/index.php/tpbci/article/view/4/13>>. Acesso em: 27 jul. 2015.

GOMES, H. E. Tendências da pesquisa em organização do conhecimento. *Tendências da Pesquisa Brasileira em Ciência da Informação*, v. 2, n. 1. p.60-88. 2009. Disponível em: <<http://inseer.ibict.br/ancib/index.php/tpbci/issue/view/2/showToc>>. Acesso em: 1 ago. 2015

GUIMARÃES, J. A. C. A dimensão teórica do tratamento temático da informação e suas interlocuções com o universo científico da International Society for Knowledge Organization (ISKO). *Revista Ibero-Americana de Ciência da Informação*, v.1, p.77-99, 2008. Disponível em: <<http://periodicos.unb.br/index.php/RICI/article/view/2761>>. Acesso em: 1 ago. 2015.

GUIMARÃES, J. A. C. et al. A dimensão temática da pesquisa em organização do conhecimento: uma análise de domínio dos congressos nacionais e regionais da ISKO Brasil, Espanha e América do Norte. *Scire*. v.20, n.2, p.19-25, jul.-dez. 2014. Disponível em: <<http://ibersid.eu/ojs/index.php/scire/article/view/4177>>. Acesso em: 1 ago. 2015.

HJORLAND B., ALBRECHSTEN H. An analysis of some trends in classification research. *Knowledge Organization*, v.26, n.3, p.131-139, 1999.

IBEKWE-SANJUAN, F.; SANJUAN, E. Knowledge organization in the last two decades: 1998-2008. In: GNOLI, C.; MAZZOCCHI (Ed.). *Paradigms and conceptual systems in Knowledge Organization, Advances in Knowledge Organization*, v.12, p.115-121, 2010.

LÓPEZ-HUERTAS, M. Some current research questions in the field of Knowledge organization, *Knowledge Organization*, n. 35, n.2-3, p.113-136, 2008.

MCILWAINE, I. C. Trends in knowledge organization. *Knowledge Organization*, n. 30, v.2, p.75-86. 2003.

McILWAINE, I. C., WILLIAMSON N.J. International trends in subject analysis. *Knowledge Organization*, v. 26, n.1, p.23-29,1999.

SAUMURE, K.; SHIRI, A. Knowledge organization trends in library and information studies: a preliminary comparison of the pre- and post-web eras. *Journal of information science* v.34,

n.5, p.651-66, 2008. Disponível em: <<http://jjs.sagepub.com/content/34/5.toc>>. Acesso em: 27 jul. 2015.

SMIRAGLIA, R. About knowledge organization: an editorial. *Knowledge Organization*, v.32, n.4, p.139-140. 2005.

SMIRAGLIA, R. Whither knowledge organization? An editorial. *Knowledge Organization*, v.33, n.4, p.8-10, 2006.

SVENONIUS, E. Directions for research in indexing, classification, and cataloging. *Library Resources & Technical Services*, v. 25, n.1, p. 88- 103. January/March, 1991. Disponível em: <<http://downloads.alcts.ala.org/lrts/lrtsv25no1.pdf>>. Acesso em: 25 jul. 2015.

ISKO-Brazil and the research groups in knowledge organization



Evelyn Orrico

Federal University of the State of Rio de Janeiro
evelynorrico@unirio.br

1 Introduction

The role of Knowledge Organization is fundamental, once through KO we organize our understanding of the world. In other words, through the conception of knowledge organization one can identify how different social groups are represented over time. In order to illustrate such statement, we can recall curious information: the researcher Freire, in his weekly column in Amazon's Daily, on August 2, published a text on the incidence of suicides in indigenous communities. In his text, he takes the words of another researcher to state that the term suicide does not even exist in most Indian languages, which makes us infer that the suicidal practice either did not exist or was not recognized as an identifiable action in these communities. We can state that because it is recognized that the relationship between the things of the world and the world itself is established via language. If there is no term for a particular action that a social group carries out in the world, it is because this action does not play a worthy social role for this group, and therefore, without the need to be designated by a term specifically created for this purpose.

By contrast, actions are also designated due to the understanding that different societies in different historical moments make of them. Also for illustration, we can mention the request made to the publisher Editora Melhoramentos, which edits Michaelis Dictionary through the website Change.org, to change the definition of the

entry *marriage*, published by Senra (2015). Until then, the concepts of this entry were defined as 1. legitimate union between man and woman.

The online petition presented by Eduardo Santarelo was justified by the social changes that have taken place in recent decades and that allow people in homo-affective relationships to establish legally and judicially recognized homo-affective unions.

What these two examples show us is that knowledge organization and information representation is a field of knowledge that demands frequent and contextualized reflection, given the constant transformations social relationships go through. It would then be important to acknowledge the impact an association of knowledge organization has caused to the strengthening of the field itself.

2 A panorama of ISKO-Brazil

The Brazilian Chapter of ISKO, whose creation arose from the researchers participating in the Work Group 2 of ANCIB, Knowledge Organization and Representation, had been organized by the group since 2005 and finally in 2007, during the 8th ENANCIB (Brazilian Association for Information Science Annual Meeting), ISKO-Brazil was finally installed and had its regulations approved.

From its creation to the year of 2015, the curve of associate members is growing, as shown in Chart 1, data from 2000.



Chart 1 - Associate members 2009-2015 period

Regarding the category of members, there are three possibilities: active professional, retired professional and student. And the last four years present the following distribution, shown in Figure 2.

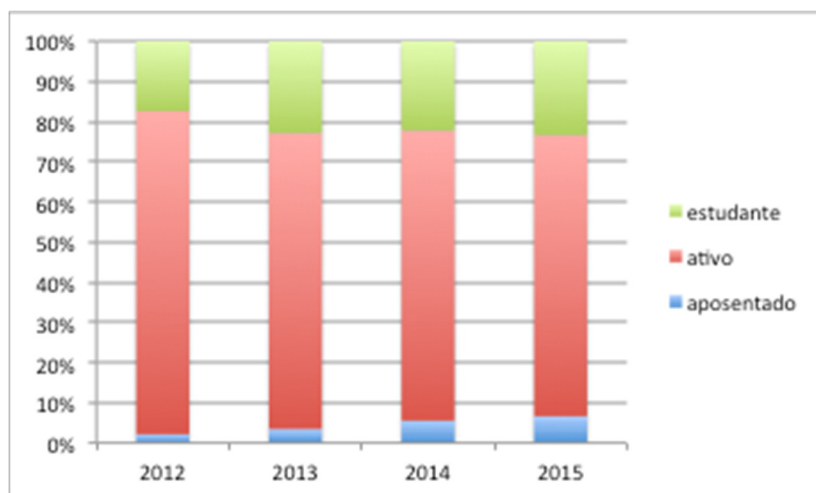


Chart 2 - Category of members 2012-2015 period

The trajectory of the association may be regarded successful by analyzing the curve of associate members, which has maintained upward. It is also noteworthy that the balance between those leaving the active professional life due to retirement and the ones entering in the professional life, as young students, has remained in balance. Although the number of retirees has increased, the number of students has increased as well. I call attention to this fact because the presence of young professionals may reflect positively on the continuity of the Association. Considering that most of our associates comprises professionals who are active, I try to reflect on the impact that this body of members, and hence ISKO-Brazil, might be causing on the scientific production of the field itself.

3 Impact Perspective

The path taken to visualize the impact ISKO Brazil may be causing on the academic production of the field was to review the CNPq (Brazilian National Research Council) research groups, conducting a search by the terms *knowledge organization and representation* related to our area of interest and with our researchers. The results are from August 2015, therefore recent, and this review presented data that allow us

to think about future actions. First, we searched the keyword *knowledge organization* without any kind of filter and without requiring the exact words. Such search has retrieved 101 groups registered in CNPq from various areas of knowledge: Health Sciences (2); Exact and Earth Sciences (32); Humanities (18); Applied Social Sciences (38); Engineering (3); Linguistics, Language and Arts (8). What is clear from this result is that these terms instigate research in several conceptual frameworks.

However, when we searched the term *knowledge organization*, filtering by exact words, this number dropped to 25 groups, with the following distribution by areas: Exact and Earth Sciences (2); Applied Social Sciences (23). This result confirms a theoretical and conceptual framework that inserts research on knowledge organization in that list of subjects that belong to the area of Social Sciences, which include Librarianship, Archival Sciences, Information Science.

Reflecting particularly on the relationship between ISKO-Brazil and creating research groups dedicated to the theme of knowledge organization, we refined the review in order to identify the period of time in which the groups were created. The CNPq Directory allows the following search possibilities for the creation of research groups: less than 1 year; between 1 and 4 years; between 5 and 9 years; between 10 and 14 years; and 15 years or more. The results include research groups updated in the last year and groups considered not updated, i.e., those who did not suffer update in the last 12 months.

We found the following data, arranged in table 1:

Table 1 - Period of research groups' creation and their members

Período de criação	Quantidade de grupos de pesquisa	Grupos com presença de associados	Líder associado	2º líder associado
Menos de 1 ano	0			
1 a 4 anos	10	5	4	1
5 a 9 anos	7		4	3

10 a 14 anos	4	2
15 anos ou mais	4	

In the groups with 1-4 years of creation, out of the 10 existing groups, five of them have ISKO members, four as leaders and one as second leader, in which the first is not associated.

In the groups with 5-9 years, out of the seven groups created, 3 of them have the second leaders as members, and 4 of them have associated leaders, with the following distribution: 2 groups have the 1st and 2nd leaders associated; and in 2 groups only the second leaders are members. In the four groups created between 10 and 14 years, two of them have their first leaders associated.

Out of the four groups set up 15 years ago or more, three present associated leaders and one of them, in addition to the leader, the second leader is also part of ISKO-Brazil. Considering the eight years of ISKO Brazil establishment, the sets of groups that could have been impacted the most, would be the top three of our list: between 5 and 9 years; between 1 and 4 years and less than one year.

Except for this last year, when groups were not created on the theme, the two sets, between 1 and 4 years and between 5 and 9 years, were the ones presenting major growth, indicating that professionals of similar interest aggregated efforts for a common object. Therefore we suggest trying to understand this data given through association as a possible mediator of this impact.

4 Association

We reflect on the formation of associations in order to understand, or possibly propose strengthening policies for ISKO-Brazil, given that the proposal thought by Tocqueville (1998) can contribute to this type of organization. The presumption of this form of organization is the participation of all members in decision-making, engendering a creative way to jointly pursue the object of their shared interests.

The author makes his reflections grounded on the American society and its associative movement, highlighting that this movement results from people gathering in groups under a common interest from a formal and institutionalized sociability with

some political power. He emphasizes that such practice is of collaboration and solidarity.

This leads us to think as Fonseca (2008) when he studied associations with legal personality in Rio de Janeiro between 1903 and 1916. The author pointed out that the members, gathered around the same interests, not only converge efforts, but also give visibility to their struggle by becoming stronger interlocutors in the debate with the rest of society and the state, which certainly increases the possibilities of changing social reality.

The main objective of ISKO Brazil is to promote dialogue among its members, not indeed to establish relationship with the state, however, the association can contribute to represent social practices and thereby help to settle the transformation of social reality. The dialogue among the members is evidenced on the three objectives of the Brazilian chapter of ISKO (<http://isko-brasil.org.br>) which, in our view, coincide with the general presumptions of the association, which is converging common interests and join forces to solve problems:

- Promote research, development and application of conceptual knowledge organization systems that promotes the study of philosophical and semantic aspects of knowledge structure;
- Provide means of communication and networks in knowledge organization for its members; and,
- Act as network interface between national and international institutions working on issues related to conceptual organization and dynamics of knowledge.

Also according to Fonseca (2008), associations saw their publications as effective ways of achieving their members. Silva (2013) even stated that these publications are privileged sources that feedback their own constitution.

In the case studied by Silva (2013), association converged to institutionalization of the archival field, in the case of ISKO Brazil, the creation of the association aimed at gathering researchers together, specifically around a wide and multifaceted theme, which is Knowledge Organization.

5 Discussion e conclusion

The initial data collection involving the groups recorded in CNPq database showed that in almost all areas of knowledge the research interests pervades the terms

organization and knowledge, but that may be trampled in various conceptual frameworks, as shown by the inclusion of 101 groups retrieved in the search: Health Sciences (2); Exact and Earth Sciences (32); Humanities (18); Applied Social Sciences (38); Engineering (3); Linguistics, Language and Arts (8).

However, when the search suffers the restriction of the filter "search by exact word", that number drops to 25 and the Knowledge areas are also restricted to two - Exact and Earth Sciences and Applied Social Sciences - the latter being the one concentrating most groups, pointing to the conceptual specificity of the phrase knowledge organization.

Considering the association's premises and reflecting on the data we found, we realize that the creation of the Brazilian chapter of ISKO encouraged and favored the meeting of researchers towards the deepening of Knowledge Organization issues. This can be explained by the growing number of research groups registered in CNPq in the years following the date of establishment.

On the other hand, the discreet presence of students in the Association, although in slightly upward curve, called our attention. Their presence would be the most effective indicative of a renewing power - and at the same time fomenting - renewal of the field, ensuring the prospect of greater vitality.

Another fact worth noting is the lack of establishment of research groups in this field in the last twelve months. One possible explanation is that the groups created for less than nine years, a period that includes the creation of the Association, are now beginning to reap the rewards of maturity of new members formed within the older groups. We will wait for the next few years to verify this possibility - or not.

As the carried out search aimed to identify the formed groups, it is necessary, from now on to refine searches, identifying, for example, among those groups which ones are updated, productive and configured.

It is also noted that the older groups, although in small number - only four - three of them have their leaders as members, and in one of them the second leader is a member. They are groups created by researchers that had traditionally already been working in the area and until now continue to produce in the field of Knowledge Organization; some of them having participated in the meeting that expressed "strong interest in creating the Brazilian Chapter of ISKO," according to the minutes published on the website <http://isko-brasil.org.br>, in the tab Memory.

We all know that the volunteer work of those involved in proposals of associative nature needs to be in constant assessment and regular reviews on the adopted routes must be promoted. This table on Perspectives in Knowledge Organization gave us the opportunity to reflect on our own associative path ISKO-Brazil has taken in these eight years of effective existence.

References

The references was made following the ABNT rules.

FONSECA, Vitor Manoel Marques da. *No gozo dos direitos civis: associativismo no Rio de Janeiro, 1903-1916*. Rio de Janeiro: Arquivo Nacional, Niterói: Muiraquitã, 2008.

FREIRE, José Ribamar Bessa. Como matar índio com giz e apagador. *Diário do Amazonas*, 2 de agosto de 2015. <http://www.taquiprati.com.br/cronica.php?ident=1155ABNT6023e10520>. Acessado em 2/8/2015.

SENRA, Ricardo. *Dicionário Michaelis muda definição de casamento após pressão online* - Da BBC Brasil em Londres, 2 julho 2015. <http://www.bbc.com/portuguese/noticias/2015/07/150702salasocialdicionariomichaeliscasamentors>. Acessado em 2 de agosto de 2015

SILVA, Eliezer Pires. *Memória e discurso do movimento associativo na institucionalização do campo arquivístico no Brasil (1971-1978)*. 2013. Tese de Doutorado (Doutorado em Memória Social). Programa de Pós-Graduação em Memória Social, UNIRIO, Rio de Janeiro, 2013.

TOCQUEVILLE, Alexis de. *A democracia na América*. 4. ed. Belo Horizonte: Itatiaia, 1998.

Research perspectives on knowledge organization: reflections from PPGCI IBICT-UFRJ academic papers



Rosali Fernandez de Souza

Brazilian Institute for Information in Science and Technology
rosali@ibict.br

Introduction

This text was written for the presentation at the Round Table "Research Perspectives on Knowledge Organization in Brazil" at the Third ISKO Brazil meeting, held in Marilia/SP, in 2015.

Aiming at bringing a reflection on the main issues of the meeting on Knowledge Organization, namely:

What is the role of Knowledge Organization and its social commitment in relation to the challenges of cultural diversity? What part does the contemporary subject play in the multitude of actions required for the organization of knowledge? What are the epistemological, technological and social perspectives put for teaching, research and practice in knowledge organization to address a world permeated by plural culture? How has the field of knowledge organization been built within the scientific universe? (http://isko-brasil.org.br/?page_id=21 access on 30 June 2015).

I considered that this would be an important opportunity for myself to reflect as an advisor and researcher, and at the same time, to inform the community about the academic-scientific production of IBICT-UFRJ Post-Graduation Program in Information Science in Knowledge Organization as a contribution to research perspectives on the questions proposed for discussion at the meeting.

1 Knowledge Organization and its social commitment in relation to the challenges of cultural diversity

In the first question: What is the role of the Knowledge Organization and its social commitment to the challenges of cultural diversity? I mention the student Rodrigo Piquet Saboia de Mello's Master's dissertation, advised by myself, completed in 2015. The title: A classifying look on the imagery collection of Darcy Ribeiro's scientific expeditions to Urubu- Kaapor Indians in the Indian Museum in identifying elements of indigenous culture and Brazilian ethnic and key cultural diversity (MELLO, 2015) - an extensive title, but it explicits the elements of the research problem.

Having the classification as the focus that permeates the research, he presented, as the analytical corpus, the imagery collection of photos, archived on the Indian Museum (RJ), of the scientific expeditions to the Urubu-Kaapor Indians who counted with the anthropologist Darcy Ribeiro - when ethnologist of Studies Section (SS) of the Indian Protection Service (IPS), official indigenous agency of the Brazilian government.

We can consider that, at least five factors are important to highlight observing the corpus of the collection of photographs in question: the view of the anthropologist, the view of the photographer, the view of the information professional, the view of the information system of the Indian Museum, where the collection is archived, and the view of the user. These different views necessarily reflect different viewpoints that influenced the idealization of the photo to anthropological and cultural objectives and institutional goals, in the definition of the photo frame as the art of photography and the identification and selection of the elements for the thematic and technical indexation. That is, they are all aspects to be taken into consideration in the processes of representation and organization of photographic items of expeditions to the search needs, access and retrieval for potential users, as well as to preserve the collection of the Indian Museum.

The analysis of the photos evidenced important elements of the indigenous culture of the Urubu-Kaapor Indians such as: Food, Disease, Feast of the Nomination and Featherwork Art. As for the Brazilian ethnic and cultural diversity, examples of categorical themes were identified: Oratory, Negros, Country and Bumbameu-Boi Party.

Another thematic category identified as relevant to the purposes of photos indexation of the expeditions was the Cordial Man - concept created by Darcy Ribeiro. In the analysis of the photos in this category, it was represented by the two key personalities who participated in the expeditions: the English researcher, at the

time doctoral student at Oxford, Francis Huxley during the Second Scientific Expedition and Darcy Ribeiro himself, coordinator of two expeditions.

As for the social anthropologist Francis Huxley, the title of his work - product of the expedition - is noteworthy, entitled *Affable Savages* reveals two features captured from the Urubu Kaapor Indians: they are wild but at the same time affable. Moreover, it is also important to highlight the interest of this research to investigate the kinship system of the Indians.

Therefore, it is important to report a comment the student made during our discussions that illustrates, in my view, the social commitment of knowledge organization regarding the challenges of cultural diversity that coincides with the interest shown by Francis Huxley. The student reported (as an Indian Museum official) that in conversation with an Indian member of the Urubu-Kaapor tribe, when showing the collection of photos of expeditions in the catalogue, and when exemplifying the indexing process, he randomly took a particular photographic item of the Museum collection. When seeing such a photo, the Indian would have, with emotion, identified his grandfather, chief of the tribe - a look not foreseen by the anthropologist, by the photographer or the information professional - that is, the subtlety of a look from the simply human point of view, which reflects an aspect of the social and cultural organization of the tribe.

The anthropologist, ethnologist Darcy Ribeiro was photographed at various moments, and in photographic series, in images that suggest familiarity and proximity to the Urubu-Kaapor Indians, depicting a cordial and friendly relationship with the Indians - objective that meets the specific aims of the expeditions and which also indicates the relevance of this thematic category.

Kossoy's contribution, cited in Mello's work, on the iconographic analysis:

The iconographic analysis in the case of photographic representation is located halfway to the search for content meaning; to see, to describe and to find is not enough. This is the moment of a deep incursion into the represented scene, which is only possible if the visual fragment is understood in its interiority. Therefore, it is necessary, along with solid knowledge on the depicted historical moment, to reflect focused on content, however, on a level beyond that which is given by the iconographic see. This is the deepest stage of research, whose boundaries are not crystal defined. Often the researcher is surprised reflecting on this post-iconographic plan, searching for links to understand the life that was. (KOSSOY, 1989, p. 65 apud MELLO, 2015, p. 90)

He illustrated the reflections I made, opening to the perspectives in knowledge organization research.

2 The contemporary subject in the multitude of actions for knowledge organization

The research that made me reflect on the participation of the contemporary subject in the multitude of actions required for knowledge organization was the doctoral thesis by Fernando Ewerton, advised by Geraldo Nunes (UFRJ) and co-advised by myself, completed in 2013, entitled Political classification and journalistic discourse: how transnational news agencies represent candidates in 1989 and 2010 Brazilian presidential elections (FERNANDEZ JUNIOR, 2013).

The material studied consisted of a content analysis of 305 journalistic texts distributed by four transnational news agencies during the Brazilian presidential elections of 1989 and 2010. The analyzed agencies were: the French agency Agence France Presse (AFP), the American agency Associated Press (AP), the Spanish agency EFE, and the British/Canadian agency Reuters. The objective of the research sought to identify relevance criteria applied by transnational news agencies in the representation of candidates in two important moments in national politics.

As Fernandez Junior comments:

The representation of information is common concern for journalists and information scientists, and among other areas and professions. They work to provide texts/documents that satisfy the reader/user in their demand for information. And as the understanding of this demand becomes more complex, with the proliferation of digital platforms, challenges for those who produce and for those who index and retrieve information increase, which justifies a closer relationship between the two fields of study. (FERNANDEZ JUNIOR, 2013, p. 27).

The above indicates perspectives for research on knowledge organization in interrelated issues involving Information and Communication sciences.

In this study, language, meaning and representation are inextricably linked. In the words of Orrico and Oliveira (2006, p.155), cited by Fernandez Junior, "things do not mean in themselves, but enunciators / users construct the meaning", from "practices and symbolic processes through which representation, meaning and language operate "and which,

In this conception, it is assumed that the social actors use the conceptual framework of their culture to build representational meaning, in this case, the language. Through this system social actors make the world meaningful and through it they communicate this world to other actors. (ORRICO; OLIVEIRA, 2006, p. 155 apud FERNANDEZ JUNIOR, 2013, p. 202).

In this sense, the methodological procedure, which is closest to the interest of this discussion, was the creation of a classification of types of terms used in newspaper articles when characterizing the candidates:

- Political Term: related to the candidate and/or the political party;
- Partisan Term: which relates the candidate to the name of the party;
- Personal Term: related to the person, not applied to the party;
- Functional Term: related to activities, roles and functions of the candidate;
- Relational term: in relation to other candidates and parties.

The creation of this political classification to represent the candidates for Brazilian presidential elections in the 1989 and 2010 elections was important, even essential, for the analysis of the terms in journalistic texts used by news agencies and among agencies in both elections, searching for similarities and particularities in the representation of candidates.

In the concluding words of the thesis, the doctorate student highlights that:

[...] the view from Information Science perspective offers a particular approach to understand the relevance criteria that lead to the predominance of certain terms/concepts in journalistic texts, which can be exploited in other themes and discourses in this and other fields of study, helping to understand the representations circulating in the digital society (FERNANDEZ JUNIOR, 2013, p. 205).

Which suggests perspectives in Knowledge Organization and Representation research.

The continuation of the research in the same line of research in post doctoral stage aimed to broaden and deepen the reflection on the relevance criteria applied in the construction of journalistic discourse in the international context. The research, entitled Relevance in Journalistic Discourse: organizational, idiomatic and cultural aspects in the representation of the Brazilian presidential candidates in the 2010 election (FERNANDEZ JUNIOR, 2015), was advised by me and completed in 2015.

The post-doctoral research investigated the issue of relevance of journalistic discourse, using as methodology, the relevant parameters traditionally used in information retrieval systems. As objectives of the research, Fernandez Junior points out:

[...] assess whether **the relevant criteria of the agencies are stable while other aspects** (language) **change**; whether, agencies, **different vehicles use similar criteria to assign relevance**, applying **different weights and multiple criteria, used interactively**; whether **different political classifications, in different languages, use similar terms but apply different weights**; and whether **there is consistency of criteria among agencies and vehicles of common national, idiomatic or cultural origin**. (FERNANDEZ JUNIOR, 2015, p. 6, bolded by author)

As a contribution, Fernandez Junior highlights that

[...] when focusing on relevance criteria and political classifications present in the journalistic discourse on international news websites, we seek to contribute to the understanding of the representations found in texts stored in digital databases, at a time when the internet plays a central role in distributing information on a global scale (FERNANDEZ JUNIOR, 2015, p. 54)

Therefore, opening new perspectives for research in Knowledge Organization in the context of the digital society.

3 Epistemological, technological and social perspectives for teaching, research and practices in knowledge organization in a world permeated by a plural culture

In characterizing this session, I include Thulio Pereira Dias Gomes Master's dissertation, advised by me and co-advised by Sgarbi Goulart (UERJ), concluded in 2015, entitled: Cartoon is the subject: documentary analysis of cartoons (GOMES, 2015).

In this investigation, I highlight two questions as important to this discussion: to consider and to analyze the Cartoon as a document and to investigate possible correlations between the information processing procedures in Documentary Analysis applied to Cartoon.

Due to inaccuracies in the definitions of mood drawings, this investigation addressed the concepts of Cartoon, Comic Strip and Caricature to identify peculiarities

and specificities of cultural aspects, observed according to etymological aspects and definitions found in language dictionaries - Portuguese, English and French and other sources.

Cartoon was defined as a "discursive genre of hybrid use of imagery and verbal languages, characterized by temporality and marked by satire and critics regarding a particular event, often of political nature" (GOMES, 2013, p. 26-27).

In the investigation of the Cartoon in the pragmatic dimension of information, the variables of the documentary analysis overall process proposed by Lara (Lara, 1993) were analyzed, adapted to the context of cartoon documentary analysis. As global process variables of Cartoon documentary analysis, we identified: the descriptive structure of cartoon, gross information on the cartoon, the knowledge domains interested in the cartoon, the cartoon custodial institutions, the cartoon users and the knowledge background of the analyst and his/her ideological formation (GOMES, 2015 p.134).

As a conclusive aspect, Gomes highlights:

The existence of documentary analysis variables is evidence that this process of knowledge organization is historical, i.e., confined to specific social conditions. Consequently, documentary representation, whatever the document, cannot be assumed to be neutral and impartial, which allows us to infer the impossibility of a universal and timeless proposal for documentary analysis of cartoons (GOMES, 2015, p. 134)

Which not only evidences indispensable aspects to be considered when treating the documentary analysis of cartoons, but also increases the prospects of research in information representation and knowledge organization in different types of documents.

4 Knowledge Organization domain in the scientific universe

Considering the knowledge organization domain in the scientific universe, three studies can be mentioned in research areas and fields of knowledge in exact sciences, agricultural sciences and humanities.

In the area of exact sciences, I mention the Luana Farias Sales Marques' doctoral thesis entitled Semantic Integration of Scientific Publications and Research

Data: a model proposal for enhanced publications for nuclear sciences, oriented by me, and having Luis Fernando Sayão (CNEN) as co-supervisor, concluded in 2014.

The concept of Enhanced Publication, correlating scientific publications and research data requires a semantic integration based on a consistent Knowledge Organization and Representation of Information. In the e/Science world this work is possible through technological resources and the development of Semantic Web standards, on the basis of consistent conceptual relationships.

Nuclear Sciences is an area of multi and interdisciplinary amplitude of technological scientific, political and social scope, and which has an intensive generation of scientific data from all types, of experimental nature and generated by simulation.

As a conclusive aspect, Marques highlights that:

[...] despite the strong presence of other disciplines in the study of new concepts of scientific publications on the environment of eScience - which has technology as a determining factor, especially from Computer Science - theories, methodologies and practices postulated by Information Science and Library prevail as a structural component, and are strongly underlying all the studies and applications of the area, especially when they involve issues of representation and meaning (MARQUES, 2014, p. 234)

Which are typical issues of knowledge organization

For the present discussion, it is also important to point out the Sales' words:

[...] the feasibility of incorporating principles and theories of Library and Information Science for the technical and scientific knowledge organization in the world of eScience (MARQUES, 2014)

Which indicates the possibility of expanding knowledge organization in the role of semantic integration in the development of taxonomies and ontologies within eScience - a reality that we are already experiencing.

The great area of Agricultural Sciences is represented, in this discussion, by the dissertation entitled Agriculture Domain Representation in the Context of Knowledge Organization by Leandra Pereira de Oliveira, advised by me and concluded in 2015. The central focus of the research was directed to knowledge organization systems (KOS) as a source for conceptualization and delimitation of a knowledge domain.

The abstract of OLIVEIRA's dissertation presents the nature of the research, the objectives, the methodology, the findings and contribution of the study to the domain of knowledge organization in the scientific world:

This work is an exploratory research that has as main objective to investigate representational models, aiming to obtain particularities of Agriculture as a knowledge domain. Bibliographic and documentary search were done in order to achieve the specific objectives to identify the etymological origin and the historical aspects of Agriculture, to investigate the area of Agriculture in national and international classifications of knowledge, and to observe another reference model application within an educational context. The classificatory instruments analysed were: the FAO Categorization scheme AGRIS/CARIS, the CNPq's Tabela de Áreas de Conhecimento, and the representation of the Brazilian post-graduate programs in 'Agricultural Sciences' in CAPES classification. The analyses revealed the complexity surrounding the representation of a knowledge domain. In the case of Agriculture, the difficulty lies in the plurality of objects of study that transcend disciplinary and interdisciplinary domains. This feature, on one hand, motivates and enriches the scientific and technological development. On the other hand it leads to difficulties in the representation of Agriculture referential and representational knowledge organization models. We conclude that the analysis of different representational models is a valid approach to grasp different visions for the area reflected by the organization and representation of knowledge as practiced by communities of discourse.

The conclusions of the study show the importance of research in knowledge organization always be attentive to the different looks of scientific knowledge by discursive communities.

The great area of Human Sciences is represented by the doctoral thesis: The process of scientific communication in the field of Defense in Brazil: from the generation of knowledge to the provision of Information by Jaqueline Santos Barradas, advised by Lena Vania Ribeiro Pinheiro (IBICT), concluded in 2015, when I participated as a member of the Examining Board.

This paper, although focusing on scientific communication process from the generation to the availability of information, interests us for the present discussion as it showed not only the difficulty and complexity of defining and naming Defense as a field of knowledge, but also the consequences of decisions taken in this direction for the analyzes and subsequent interpretation of the results and conclusions.

An important finding of the research for the present discussion in the words of Barradas, author of the thesis:

Another finding of this research was to verify that there is no consensus on the concept of Defense, not even on which area (of knowledge) it must be hierarchically linked. There is still much to delimit, highlight, construct: its origins and frontiers in Brazil and abroad: field of knowledge in which it is inserted; epistemic and interdisciplinary intersection with other areas - such as strategic studies, military science, Political Science and International Relations [...] (BARRADAS, 2015, p 183).

Indicates the delicate role in the delimitation decisions of the scope of the areas, fields and domains of knowledge depicted in research on knowledge organization.

5 Conclusions

From the PPGCI IBICT-UFRJ academic papers reported, some points emerged as research perspectives on knowledge organization in Brazil.

The Brazilian ethnic and cultural diversity and its historical, scientific and social documentation of activities produced in the form of library collections of documents, museums and archives presents rich potential and a challenge to be explored in the research of knowledge organization. The specific case of the photo collection of scientific expeditions, due to the performance of the Indian Museum, is a representative example of research of photographic, anthropological, sociological look brought by the information professional that impacts the decisions of knowledge organization and representation activities.

The analysis of newspaper articles opens research perspectives in Information Science interface with Communication Science. The development of policy classifications led to explore the validity criteria of relevance to the understanding of factors influencing the representation of information in newspaper discourses, which can be exploited in other application contexts expanding research questions in knowledge organization.

The exploitation of image as a document and as such, the object of documentary analysis, extends the focus of research in knowledge organization. In the case of mood drawing, "cartoon as the subject", the application of known methodologies not only enriched, but also prompted the development of new methods for analyzing the narratives of the images, which may be extended to other types of materials, having the theory and practice of knowledge organization and representation as central support.

The paradigm of eScience, considered the 4th scientific paradigm, is an exciting challenge for knowledge organization. The reported academic research revealed a diversity of applications of eScience aimed at integration, retrieval and sharing of information of different nature. The study also stressed the importance of hyperlinks as technological artifact of the Semantic Web and the construction of taxonomies and ontologies that are typical objects of studies and research on knowledge organization.

The papers that cover a certain area, field or domain of knowledge bring challenging perspectives for research on knowledge organization. The titles refer to: Area of Nuclear Sciences, Field of Defence and Agriculture Domain. A point to be explored is the concept of "area", "field" and "domain" of knowledge. In the three studies, the multi and interdisciplinary nature of knowledge was mentioned, which makes it difficult to define area, field or domain for different research purposes in knowledge organization.

In all studies, the important role of classification as a fundamental process of human nature and how irreplaceable the activity is in information disseminating, searching and retrieving was evidenced. In this sense, identifying research questions in knowledge organization covering different types of communication documents and media from traditional to more sophisticated and already present in the digital age becomes inexhaustible.

References

The references was made following the ABNT rules.

BARRADAS, Jaqueline Santos. *O processo de comunicação científica no campo da defesa do Brasil: da geração do conhecimento à disponibilização da informação*. Tese de Doutorado em Ciência da Informação IBICT/UFRJ/ECO. Orientadora: Lena Vania Ribeiro Pinheiro (IBICT). Co-orientador: Jorge Calvario dos Santos (UFF/Escola Superior de Guerra). Rio de Janeiro, 2015.

FERNANDEZ JUNIOR, Fernando Ewerton. *Classificação política e discurso jornalístico: como agências de notícia transnacionais representam candidatos a presidente do Brasil nas eleições de 1989 e 2010*. Tese de Doutorado em Ciência da Informação IBICT/UFRJ/FACC. Orientador: Geraldo Nunes (UFRJ). Co-orientador: Rosali Fernandez de Souza (IBICT). Rio de Janeiro, 2013.

FERNANDEZ JUNIOR, Fernando Ewerton. *Relevância no Discurso Jornalístico: aspectos organizacionais, idiomáticos e culturais na representação da informação dos candidatos a presidente do Brasil na eleição de 2010*. Pós-doutorado em Ciência da Informação IBICT/UFRJ/ECO. Supervisora: Rosali Fernandez de Souza (IBICT). Rio de Janeiro, 2015.

GOMES, Thulio Pereira Dias. *A charge é o assunto: análise documentária de Charge*. Dissertação de Mestrado em Ciência da Informação IBICT/UFRJ/ECO. Orientadora: Rosali

Fernandez de Souza (IBICT). Co-orientador: Paulo Sergio Sgarbi Goulart (UERJ). Rio de Janeiro, 2015.

GOMES, Thulio Pereira Dias. *Temas e questões em análise documental de charge*. 2013. Trabalho de Conclusão de Curso (Graduação em Biblioteconomia). Faculdade de Administração e Ciências Contábeis, Universidade Federal do Rio de Janeiro. Rio de Janeiro, 2013.

KOSSOY, Boris. *Fotografia & História*. São Paulo, Ateliê Editorial, 2012, p. 65

LARA, Marilda Lopes Ginez de. *A representação documental: em jogo a significação*. 1993. 133 f. Dissertação de Mestrado em Ciências da Comunicação. Escola de Comunicação e Artes, Universidade de São Paulo, São Paulo, 1993.

OLIVEIRA, Leandra Pereira de. *Representação do domínio da agricultura no contexto da organização do conhecimento*. Dissertação de Mestrado em Ciência da Informação IBICT/UFRJ/ECO. Orientadora: Rosali Fernandez de Souza (IBICT). Rio de Janeiro, 2014.

MARQUES, Luana Farias Sales. *Integração semântica de publicações científicas e dados de pesquisa: proposta de modelo de publicação ampliada para a área de ciências nucleares*. Tese de Doutorado em Ciência da Informação IBICT/UFRJ/ECO. Orientadora: Rosali Fernandez de Souza (IBICT). Co-orientador: Luís Fernando Sayão (CNEN). Rio de Janeiro, 2014.

MELLO, Rodrigo Piquet Saboia de. *Um olhar classificatório do acervo imagético das expedições científicas de Darcy Ribeiro aos índios Urubu-Kaapor no Museu do Índio na identificação de elementos da cultura indígena e da diversidade étnica-cultural brasileira*. Dissertação de Mestrado em Ciência da Informação IBICT/UFRJ/ECO. Orientadora: Rosali Fernandez de Souza (IBICT). Rio de Janeiro 2015.

ORRICO, E.G.D.; OLIVEIRA, C.I. de. A linguagem na construção do sentido: discurso e organização do conhecimento. In. GONZALEZ DE GÓMEZ, M.N.; ORRICO, E.G.D. (Orgs.). *Políticas de memória e informação: reflexos na organização do conhecimento*. Natal: Ed. UFRN, 2006, p.155

Search perspectives on knowledge organization in Brazil



Leilah Santiago Bufrem
São Paulo State University
santiagobufrem@gmail.com

1 Introduction and evidence of interest

The scientific literature production, a concrete element of institutionalized research practices, constitutes a dynamic process that generates and renovates the fields of knowledge, but it mainly emerges as a legitimate investigation object to the understanding of its own scientific activity field. The possibility to recognize what is produced in a certain field, as well as its distinction elements and marks, has motivated analyses aimed to reveal the complexity of socially constructed practices and to progressively discover structures and procedures of scientific investigation, providing to the researchers looks directed to the understanding of the intellectual production field in which they operate.

In the occasion of the Third Brazilian Congress on Organization and Representation of Knowledge, ISKO-Brazil (Conference 2015), the motivation for this study came from the curiosity about the configuration of the Knowledge Organization (KO) domain in the periodical literature of Information Science (IS), more specifically in relation to its dimensions: (a) epistemological, (b) applied and (c) political as suggested to capture the scope of interests of the event.

The intention to analyze how this topic is represented in the periodical literature of IS, present in the Referential Database of Periodical Articles on Information Science (Base Referencial de Artigos de Periódicos em Ciência da Informação) (Brapci), arouse from the curiosity about noticeable trends in the literature relevant to the topic of the event, in the last ten years in Brazil. It is believed that the self-knowledge about what is produced within the area and about the transformations in the research area,

whose marks permeate the scientific process, has significant support in the periodical literature.

The challenge, however, in addition to the updating on the latest scientific communications is related to the expressive production in the domain of KO, whose conceptual relations with other areas constitute the assumption of this paper. Thus, the intention also comes from the knowledge that CNPq researchers, especially the ones from the first stratum, are representatives of academic excellence, a key element in the choice of their articles as constituents of the specific corpus in this research. By including as subjects of this study the KO researchers developing a representative corpus of its production, the motivation is based on the conviction that this analysis can reveal marks of distinction in the scientific production of the area and also on the statement of its co-responsibility in the collective scientific production.

The question of how to set up the KO domain in the periodical literature of IS, represented in Brapci by researchers from CNPq begins to direct this study, in which we seek to interpret the relations that are evidenced in this corpus, trying to situate the studies in the space constituted by the dynamic articulation of the dimensions suggested by the event. The articulatory conception of these dimensions, therefore, lies in the complementarity between them, eliminating the idea that they are totally exclusive.

The choice of Brapci as a source for the retrieval of articles is justified by the recognition that this instrument allows the visualization of national products in periodical articles in the Information Science area. These, in turn, as prime channels of knowledge production and role models of a practice that is relevant to science development, present research and reflection results certified by a scientific committee in which the consultative and assistance duties support them.

2 Recognizing a domain

The scientific-communicative practices of a society are affected by their existence conditions and their legitimacy stems from the relations that producers of symbolic goods keep with each other (Bourdieu, P, Passeron, J C, 2014, p. 59). The possibility that these practices occur between the different positions of a specific cultural production field generates the diversity in modes of action in a pursuit of legitimacy in that field. To understand them it is suggested a double reading: of the field itself, focusing the practices legitimated by it and the reading of the political field

that analyzes the dominant theoretical streams. Here, the field is configured, according to Bourdieu's conception (1976), as a symbolic space in which the struggles of agents determine, validate and legitimate representations, that is, an objective space of engaged scientific commitments.

But the notion of field, which is valuable for the understanding of positions and relational movements among researchers, institutions, and thematic configurations, is not only used as a resource for the improvement of the scientific criticism and its evaluation. This is because the results of analysis in the scientific field facilitate the theoretical understanding of investigation objects and also the understanding of the notion of domain, which according to Lloyd (1995, p. 25), is necessary to delineate the formation process due to the scientific research production. This type of analysis facilitates the determination of theoretical coherence and the adequacy of procedures to the knowledge objects that are constructed to the extent that the scholars contribute to their insertion in a particular scientific field.

Therefore, the current consideration about the relations between the thematic institutionalization of a KO event and the actual scientific production that it integrates expands the understanding possibilities of this specific domain. The intention to analyze the theme selected for the event came from the curiosity about its trends in the Brazilian periodical literature. While recognizing that there will always be deficiencies in the knowledge explicitness due to new information and process of social changes, Lloyd (1995, p. 18) argues that the methodologies and the general concepts may prove susceptible to continuous improvement as a result of the analytical thinking, including the examination of conceptual systems, investigation and reasoning logics used by certain sciences and how some of them advance more than others.

As a conceptual foundation of this study, intended to analyze the KO domain in the literature, it is highlighted Esteban Navarro and Garcia Marco position (1995, p. 149, our translation) which is considered relevant to the views displayed here:

[...] discipline devoted to the study and development of the foundations and techniques of planning, constructing, management, use and evaluation of systems of description, cataloging, ordering, classifying, storing, communication and document retrieval created by the men to testify, preserve and transmit their knowledge and their actions, from their content, in order to ensure its conversion into information able to generate new knowledge.

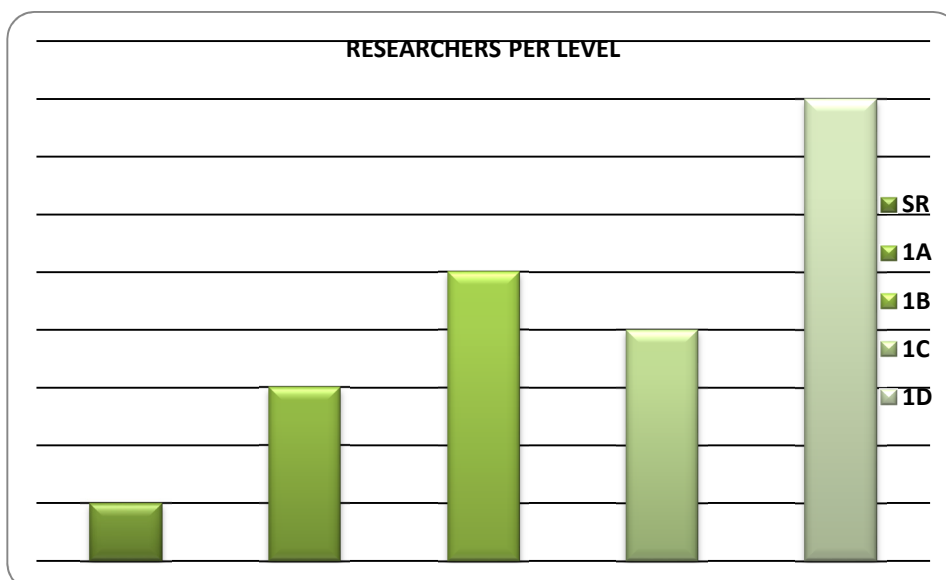
The scope of this definition, involving the aspects and the relations that compose KO, approaches Hjørland position (2008, p. 86), referring to the KO domain, as the study of the nature and quality of processes, and KO systems related to activities such as description of documents, indexing and classification in libraries, bibliographic databases, files as well as other types of institutional memory.

Hence, there is the suggestion to relate it to the representative categories of this domain, as foreseen at the Congress Isko menu. In line with this conception, Barité (2001, p. 38) highlights the theoretical benefits that KO provides to themes “related to the processing of information and in particular, the thematic processing of information; and in a less specific way - but not less important – the management of the social use of information.

It is also noteworthy that because it is an analysis of the literature in the field, this study was supported by Guimarães research (2008) on the construction of the theoretical/methodological framework of KO, based on three streams: (subject cataloguing), of North American influence, (indexing), of English influence, and the documentary analysis (analyze documentaire), of French influence.

3 Methodological path

The **first stage** of the research was the identification of Researchers Productivity in Senior Research and 1 (RS1) levels A, B, C and D of CNPq in the Science Information area. Twenty one researchers have been identified (Graph 1) , from a search on the CNPq website, by using refining filters, such as Large Area (Applied Social Sciences), Area (Information Science), Types of Scholarships (Productivity in Research) and Level of Scholarship (Senior and RS1).

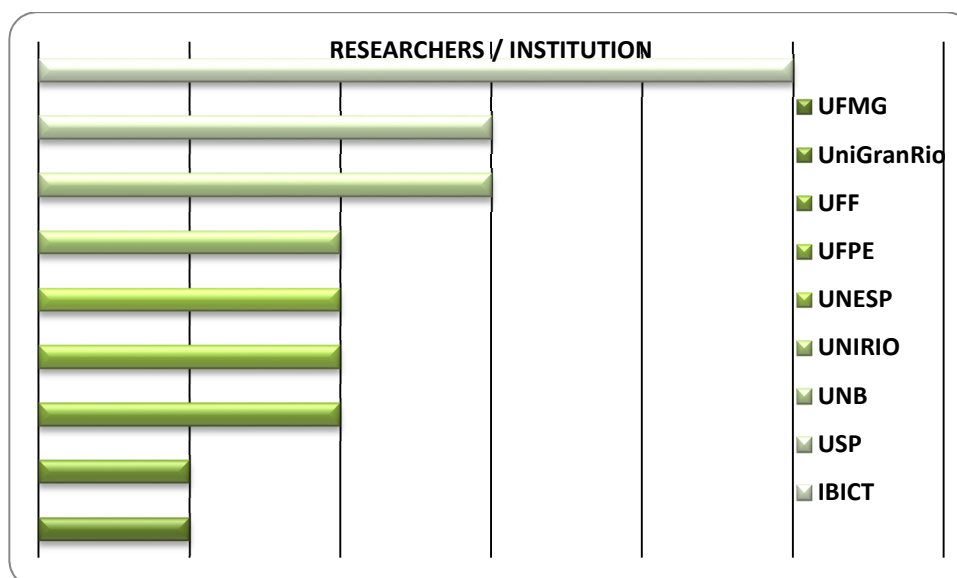


Graph 1 - Researchers category 1 and senior, from IS - CNPq
Source: CNPq website (2015)

The contextualization and the recognition of the universe and of the research subjects were facilitated by the location of the institutions in which they act, among them it is highlighted Instituto Brasileiro de Informação em Ciências e Tecnologia (IBICT) (Brazilian Institute of Information in Science and Technology) as observed in Graph 2, which also shows the higher concentration of Scholars in the Southeast. This result, related to the predominance of the region, coincides with other studies in various areas of knowledge (MENDES et al., 2006; BORTOLOZZI, GREMSKI, 2011) and it has been attributed to the highest concentration of undergraduate and graduate courses in this region, which in 2008 was blessed with 61.1% of master's and doctoral programs among all areas (SANTOS; AZEVEDO, 2009). The peculiarity of this study, however, lies in the presence of two researchers from the Northeast region in the representative picture of PQ1, which is significant compared to other regions.

The interest in this kind of relationship corresponds today to studies on social networks. Although they have not been intensively used in this exploratory study, they may contribute to the design of approaches or distances among actors that in this context may be researchers, institutions, themes, periods, or other realities represented in the networks, suggested also as a tool for the study of these relationships evidenced in the scientific production, the "Social Network Analysis" (Análise de Redes Sociais) (ARS) has been presented by Wasserman and Faust (1994) as a way to answer research questions in areas related to the social sciences,

as an appropriate analytical tool for interpretations of positions, principles, attitudes, thematic and methodological affinities, covering elements that make up the scientific conjuncture of specific domains.



Graph 2 - PQ1 researchers and senior from IS/Institution
Source: CNPq website (2015)

The characterization of the researchers required the use of some tools for extracting, compiling, viewing and organizing the data.

Tool	Purpose
Script Lattes	Automatically download Lattes curricula of research group of interest, compiling production lists, appropriately addressing duplicate and similar productions, and representing the scientific collaboration networks among researchers. Data were obtained from CNPq Lattes Platform. (MENA-CHALCO; CESAR JUNIOR, 2009)
Web Qualis	Map the quality indicators of articles published in scientific journals by field of knowledge (Social Sciences I), whose criteria are established by specialized groups of Coordination for the Improvement Higher Education Personnel (CAPES)
Vantage Point	Process, classify and represent scientific production. Its function was the creation of correlation matrices that assisted in the representation of scientific social networks.
Microsoft Excel	Create graphs and tables for the analyzed scientific production.

Chart 1 - Tools in relation to the objectives of the research
Source: by the author (2015)

Specifically, on Lattes Platform, on 6 July, 2015, there was the collection of publication data of PQ 1 and senior researchers in the period 2006 to 2015. From

these data, it was possible to reach the collaboration network shown in Graph 3, with all the PQ1 and Senior researchers.

Graph 3 - Collaboration among CNPq researchers PQ1 and Senior

Source: Lattes Platform (July 2015)

The analysis of sub-networks suggests a high degree of collaboration among the researchers from the same institution, such as the IBCTI representation in Graph 4, which lines show the degree of contribution regarding the production generated in the delimiting period of the corpus cut.

Graph 4 - Sub-network of PQ1 and Senior CNPq researchers

Source: Lattes Platform (July 2015)

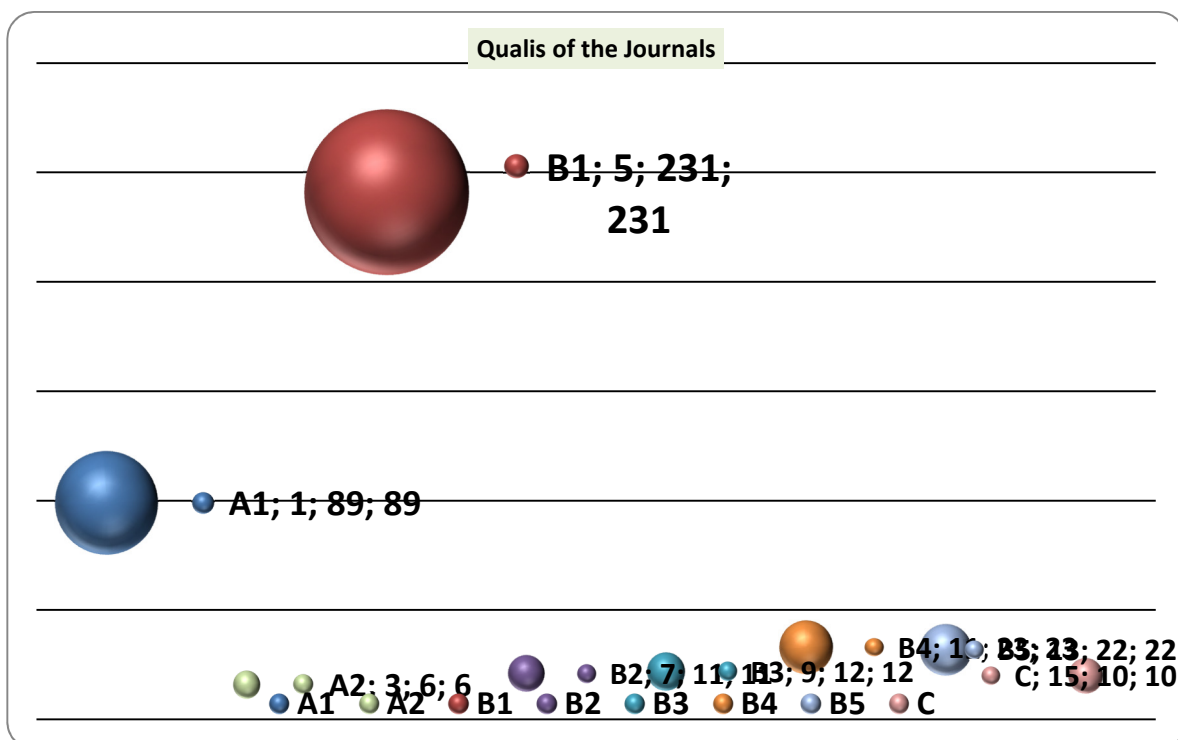
A larger and more diverse configuration of this collaboration is evidenced in Graph 5, illustrating the group of researchers from other institutions represented in the corpus with their respective degrees of intensity.

Graph 5 - Sub-network 2 of PQ1 and Senior CNPq researchers

Source: Lattes Platform (July 2015)

The articles collected on the Lattes Platform were classified according to Qualis stratum of CAPES of the periodicals in which they were published. As noted in Graph 6, there is a significant incidence of articles (231) in periodicals classified as B1, exceeding the number of other articles published in periodicals of other strata.

The high incidence of articles stratified in Qualis B3, B4, B5 and C (67) also called the attention, although they may be categorized in other areas of knowledge at a higher stratum.



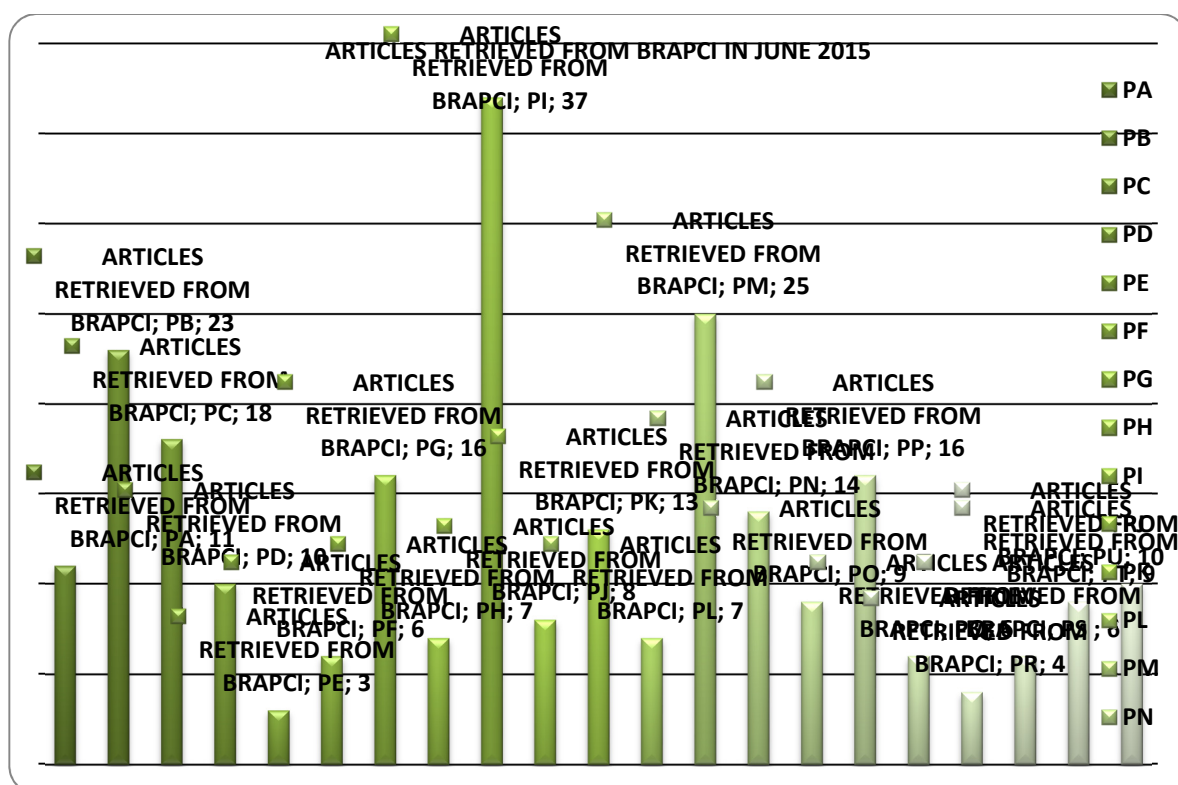
Graph 6 - Qualification of the production by CNPq PQ1 and Senior researchers

Source: WebQualis (July, 2015)

In the **second stage**, between 7 and 19 July, 2015, there was the collection of scientific articles indexed in Brapci published by RS1 Senior, A, B, C and D of CNPq. The research was conducted by the research's name in the author field, with the cut time 2005-2015, in scientific magazines.

After a survey of the works of each author, only the scientific articles were selected, eliminating other types of texts. The selective analysis determined that only the scientific articles of each author were exported to the bibliographic manager Zotero, which was used to collect, store and organize the retrieved records. The articles were organized into collections that received, each one, the name of the researcher within Zotero.

From this selection, 258 items were retrieved, and eight of them were co-authored by two researchers from the RS1 group. The inclusion of Graph 7 was motivated by the possibility of observing a picture of the authors' total production, in which the thematic configuration is wider than in Graph 8, due to the specific focus of interest in ISKO Congress, Brazil 2015.



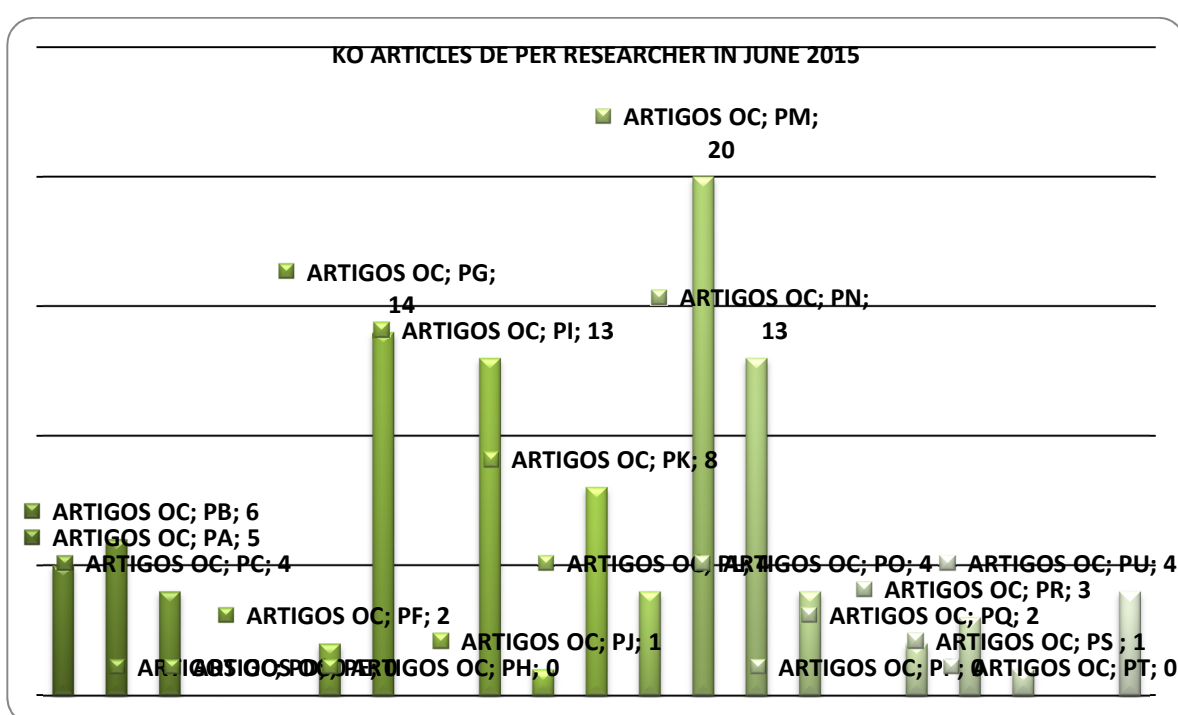
Graph 7 - Articles by PQ1 and Senior researchers, retrieved at Brapci, from 2006 to 2015

Source: Brapci (2015)

In the **third stage**, there was the reading of titles, abstracts and keywords of the 250 articles in order to identify and select as components of the corpus only those related to KO theme, totaling 104 articles, i.e. 41.6% of the whole output of RSs, which may be considered expressive. These were categorized according to the three dimensions that guide the discussions at III ISKO Brazil: 1) **the epistemological dimension** of knowledge organization, including the conceptual, historical and methodological bases of KO, as well as the interdisciplinary dialogues that are established in its scope; 2) **the applied dimension** of KO, including models and formats, instruments, productions and structures in KO and 3) **the social and political dimension** of KO, covering topics related to professional practice, ethics, culture and identity, as well as sustainable development. After being categorized, the articles were analyzed, interpreted and mapped according to the sub-categories of each dimension, as explained in the next chapter.

4. Presentation and data analysis

The 104 articles related to KO domain, in the period 2005-2015, published in scientific periodicals indexed in BRAPCI, are distributed in Graph 8, among the researchers whose names have been replaced by letters to preserve anonymity. Due to the fact that they were written in co-authorship among RS1 researchers, three of the articles present two occurrences.

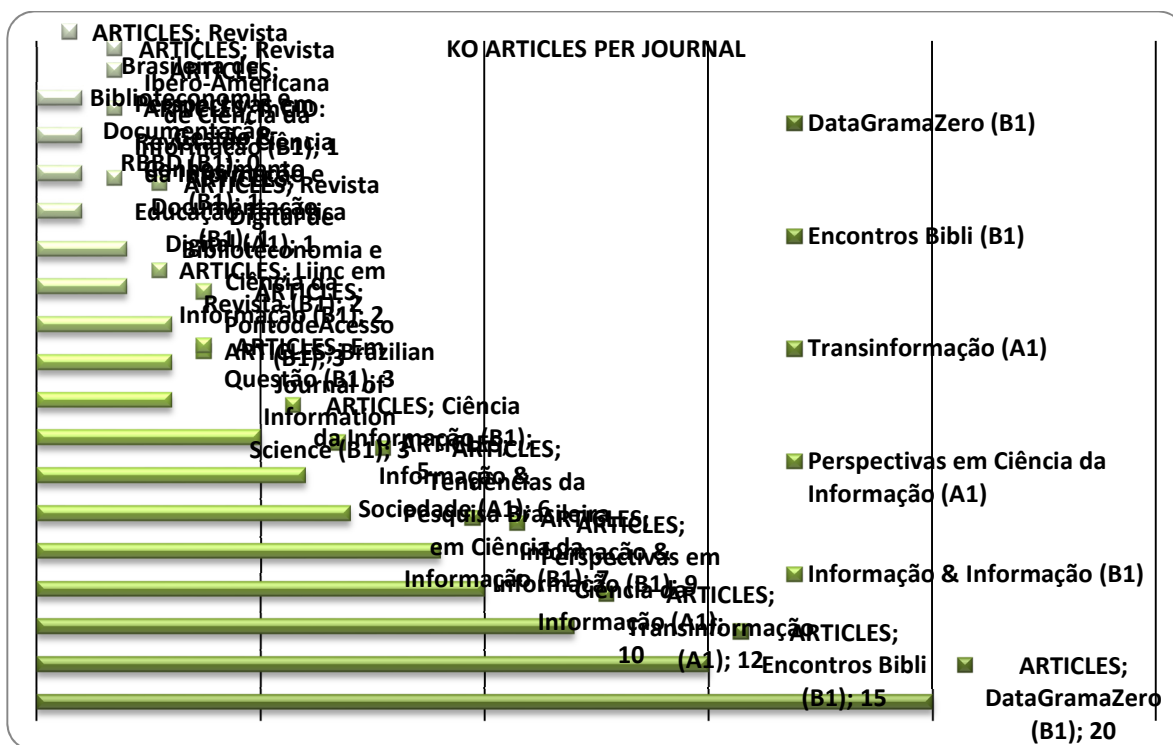


Graph 8 - Articles by PQ1 and Senior researchers, retrieved at Brapci, from 2006 to 2015

Source: by the author

The most represented periodicals in the specific corpus of KO were the magazines DataGramZero with 20 articles, Encontros Bibli with 15 articles and Transinformação with 12 articles (Graph 9). Most of them (13) are classified in the stratum B1 of Qualis CAPES, according to Graph 9. It can be noted that, although they

are not magazines with many years of existence, the first two ones have been more receptive to the themes of KO.



Graph 9 - Articles on KO, retrieved at Brapci, from 2006 to 2015

Source: by the author

The 101 articles on KO were categorized in each of the three dimensions: Epistemological, Applied, Social and Political. It is observed higher quantitative expressiveness of articles in the Applied Dimension (45 articles) while in the Epistemological Dimension 41 articles were categorized and in the Social and Political Dimension 15 articles (Graph 10).

From the content analysis, it was inferred that among the categorized articles, in the *applied dimension*, there is predominance of researches focused on models, formats, tools, products and knowledge organization structures. Among them, a specific group, the most representative in studies, is related to indexing processes, under almost exclusive responsibility of RM. The category includes articles on the indexing processes, information systems, and forms of documentary representation in indexation (model of documentary reading for scientific texts), verbal protocol and indexing evaluation.

The knowledge organization structures in this group of articles include, besides the presented and tested models, librarian tasks and tasks contextualized in libraries and in information unities, observed before the current situation that is marked by significant technological changes. These tasks are related to issues such as indexing books; documentary language; consistency of indexing; indexing in the cataloguing of university libraries; automatic indexing of articles in journals and; uses and evaluations of software.

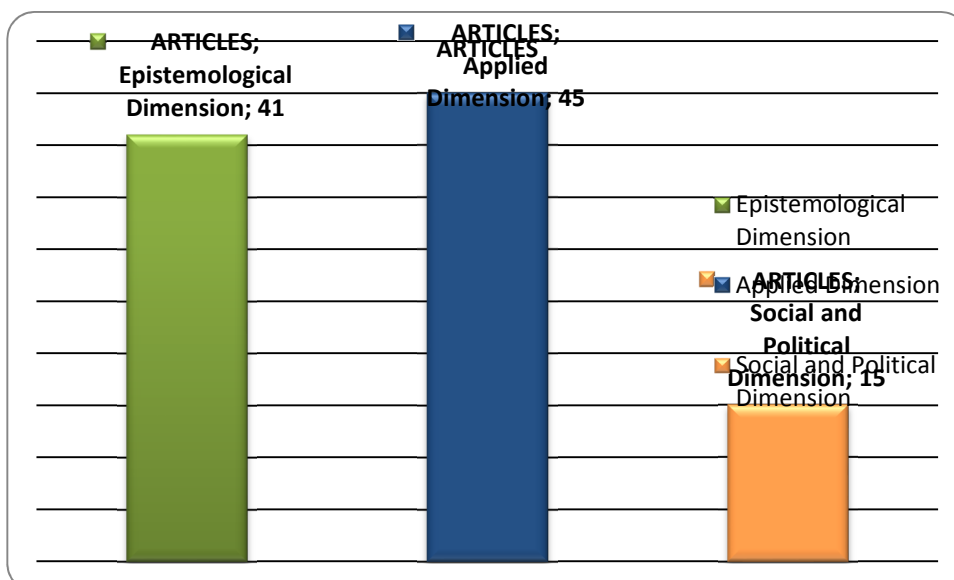
Issues related to documentary languages, interoperability and controlled vocabulary, relationships between linguistics and terminology, classification systems, and aggregation principles.

Among the products of knowledge organization, documentary languages are analyzed and assessed, including indexes and thesauri, and classificatory systems.

As subsidies of KO structures, there is the study of the following issues: evaluation methodologies; aggregation principles according to contexts and areas of knowledge; representation methodology of imagery information; literary guarantee; probative effectiveness of documents; and the practical application of content analysis and quotes for analysis and document retrieval.

The contribution to the improvement of documentary languages has been a constant in the work of researchers in RM, RN, RG and RO, whose motivations involve issues related to the assessment of languages through use, including studies on natural and controlled languages and researches on systematic indexing. Controlled vocabularies are analyzed in works such as RM, RN which are also based on quantitative assessments of documentary, alphabetical and hierarchical languages.

It is also possible to observe emerging issues related to structures and models of KO, such as the intensification of human interaction processes, the representations in the modes of literature presentation, the evaluation methods in the organization of scientific production, and the automatic creation of quotes bases, the portals and their representativeness.



Graph 10 - Dimensions

Source: by the author

The *epistemological dimension*, mainly represented by RG, RI, RK and RN, is accomplished mostly in relation to interdisciplinary dialogues in KO, a dominant theme in studies on the methodological theoretical referential in the domain. On the other hand, studies on conceptual bases of knowledge reflect a broad theoretical spectrum.

The methodological bases are represented by articles focused mainly on the relationship between the knowledge organization domain and the pedagogical practices.

The cognitive bases are the theme of articles that, in particular, address the examination of the text in the knowledge formation, as well as the dialogical process of knowing and the social context of the subjects.

The *social and political dimension*, although less represented quantitatively, exercises along with the epistemological dimension, a significant role while favoring the balance regarding the applied dimension. Expressively represented by articles that address the problematic regarding the formation and professional performance, this category had the effective participation of RM. It was backed by articles that deal with ethical issues in KO, dimension represented especially by RG.

Regarding the questions related to culture and identity in relation to KO, the same are guided in a diffuse way, relating with themes such as gender and ethics.

Final considerations

Evidence from the analysis focused on the quantitative expressiveness of occurrences start with the profile of studies on the applied dimension, highlighting the role of KO in the universe of IS and, in a way, corroborating the nuclear position of the Thematic Treatment of Information (TTI), emphasized by Guimarães (2008) and justified by its mediator nature between the production and the use of information. This “centrality” is related to the fact that only five researchers from the IS area are not part of the universe of 21 participants of the universe of Senior and **A1** researchers.

The quantitative evidence concerning the applied dimension, therefore pragmatic and focused on concrete realities, is harmonized by the epistemological dimension, also with significant representation in studies on theoretical-conceptual fundamentals, historical and technological of the domain, as well as on interdisciplinary relations.

And to prove the impossibility of compartmentalizing the views of a certain conjuncture it was evidenced in the researches members of the political-social dimension, both reflections with epistemological character and analyses on issues applied with different emphases.

Without harm to the further development of the emphasized aspects, it can be confirmed the significant diversity of the studies and the organic representation of important researchers in their field.

From the foregoing, it stands out the opportunity of content analysis of the articles, because it revealed the KO as a domain not only supported by applied studies, aimed at the needs and challenges of their daily lives, but also inspired by the social commitment, whose mission has been accomplished under the protection of principles and theories that are fundamental to the decisions regarding the ways of organizing knowledge. It was also shown significant similarities between the research recorded in journals and the original designs of the researchers thanks to references to the motivation organ in publications. On the other hand, co-authorships are noted in the diversity of this production, because they can be seen in the vast majority of the cases, combining mentees and mentors. However, this apparent dispersion related to the focus of recorded researches, when perceived in relationships that occur in the

postgraduate culture, has shown the required link between the projects of the researchers and the mentees.

References

The references was made following the ABNT rules.

BARITÉ, M. Organização del conhecimento: un nuevo marco teórico-conceitual en Bibliotecología y Documentación. In: CARRARA, K. (org.). *Educação, Universidade e Pesquisa*. Marília: Unesp-Marília-Publicações, São Paulo: Fapesp, 2001. p. 35-60.

BORTOLOZZI, F.; GREMSKI, W. Pesquisa e pós-graduação brasileira—assimetrias. *Revista Brasileira de Pós-Graduação*, v. 1, n. 2, 2011.

BOURDIEU, P. Le champ scientifique. *Actes de la Recherche en Sciences Sociales*, n. 2/3, jun. 1976, p. 88-104.

BOURDIEU, P.; PASSERON, J.-C. *A reprodução: elementos para uma teoria do sistema de ensino*. Petrópolis: Vozes, 2014.

ESTEBAN NAVARRO, M. A.; GARCÍA MARCO, F. J. Las primeras jornadas sobre organización del conocimiento: organización del conocimiento e información científica. *Scire*, v.1, n.1, p.149-157, jan./jun. 1995.

GUIMARÃES, J. A. C. A dimensão teórica do tratamento temático da informação e suas interlocuções com o universo científico da International Society for Knowledge Organization (ISKO). *Revista Ibero-americana de Ciência da Informação (RICI)*, Brasília, v. 1, n. 1, p. 77-99, 2008.

HJØRLAND, B. What is Knowledge Organization? *Knowledge Organization*, v. 35, n.2/3 2008.

LLOYD, C. *As estruturas da história*. Rio de Janeiro: Zahar, 1995. 296 p.

MENA-CHALCO, J. P.; CESAR JUNIOR, R. M. ScriptLattes: an open-source knowledge extraction system from the Lattes platform. *Journal Of The Brazilian Computer Society*, Porto Alegre, v. 4, n. 15, p.31-39, 2009. Disponível em: <<http://www.scielo.br/pdf/jbcos/v15n4/04.pdf>>. Acesso em: 5 jun. 2015.

MENDES, P. H. C. et al. Perfil dos pesquisadores bolsistas de produtividade científica em medicina no CNPq, Brasil. *Revista Brasileira de Educação Médica*, Rio de Janeiro, v. 34, n. 4, dez. 2010. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-55022010000400008&lng=en&nrm=iso>. Acesso em: 09 jun. 2015. <http://dx.doi.org/10.1590/S0100-55022010000400008>.

SANTOS, A. L. F. dos; AZEVEDO, J. M. L. de. A pós-graduação no Brasil, a pesquisa em educação e os estudos sobre a política educacional: os contornos da constituição de um campo acadêmico. *Revista Brasileira de Educação*, v. 14, n. 42, p. 535, 2009.

WASSERMAN, S.; FAUST, K. *Social network analysis: methods and applications*. Cambridge: Cambridge University Press, 1994. 857 p.

**RESEARCH IN DOCUMENTAL
ANALYSIS ON BRAZIL: THE JEAN-
CLAUDE GARDIN INFFLUENCE**

The search for efficiency in information and knowledge representation – later developments in Gardin’s thought



Johanna W. Smit
University of São Paulo
johannawsmit@gmail.com

1. Introduction

The relation between the original text and its representation followed Gardin throughout his career and he discussed the theme not only thinking about information systems, but also thinking about the various methods of text analysis.

This is not time to detail an important bibliography in which Gardin discussed the different methods of text analyses, but a quick reference to the question cannot be ignored, considering that Document Analysis (DA), in his view, represents a modality of text analyses – according to him, the most “industrial” branch of analyses (Gardin, 1974, p. 45). Emphasizing the semantic dimension of analyses, he developed a comparative view of the adopted procedures, for instance, through the content analyses carried out by sociologists and other professionals, the structural analyses of short stories and narratives, the semiological or literary analyses and the document analyses, that is, “the content expression of scientific papers in the manner of documentalists” (Gardin, 1970, p. 630, translated by the author). These comparisons led him to postulate the importance of disclosing the method of text analysis (in his formulation: how to go from T1, original, to T2, its representation), the disclosing of semantic correspondences mobilized by the analysis and the possibility of verifying the relation between the original text (T1) and its representation (T2). In other words, he proposed discussions on the analytical methods (and their explanation), about the tools (that is, the metalanguage used to extract the semantic content of the texts, according to the proposed goals) and **assessment** and **validation** strategies of the analyses.

Inserting document analysis in a larger set of text analyses certainly represented an expansion of knowledge in documentation and Librarianship by calling the attention to linguistic issues, so far solved by common sense. Between assigning a classification code to a book and discussing on the unity of analysis (word? phrase? paragraph? chapter? the entire text?), the preponderance of semantics on syntax and the automation potential of the classification process, preparation or indexation of summaries separate common sense from a more scientific and rigorous view of the library or documentary practice.

The importance of the metalanguage (ML) used to analyze texts – indexing language (IL) in Information Science – received much attention and I suppose that this issue is the most incorporated one by Information Science. A language to represent the original content of texts should be an artificial language, constructed according to goals, and therefore, distinct from natural language. The origin of the theme can be found when the archaeologist Gardin discussed how to find matches (i.e., the classification) between archaeological objects, that is, which criteria to adopt in order to gather objects in sets, distinguishing them from other sets. In the 1950s and 1960s, he developed “analytical codes” of oriental cylinders, artifacts, decorative designs, etc. always emphasizing the need of comparisons to establish relationships (Gardin, 1979, p. 87).

The concern about the importance of establishing relationships between data and information is also clear when he is part of the team discussing the conception of UNISIST in UNESCO in 1970, by stating that “it is useless to make data available, people have to be able to understand the data by establishing relationships between the data”.

It is true that the technology of the time seems to be very rudimentary today, as the selection of objects that presented common characteristics was conceived in index cards with lateral holes (peekaboo). An analysis of the concepts presented in the Koran also gave rise to the development of a structured vocabulary of concepts and tests made with these cards (Allard et al., 1963).

Investment in linguistics was very large (Gardin, 1973): let's consider that today the Group TEMMA continued this investment, but with emphasis in terminology, aiming at issues related to vocabulary control and communicational power of document analysis products.

The need for theorizing indexing languages, that is, the specific metalanguages of the environments that treat information, led Gardin to discuss their structures and composition, distinguishing in them a **lexicon**, **paradigmatic relations** and **syntagmatic relations** (Gardin, 1966, 1970, 1973); that is, whatever until that moment, many people had done based on common sense, now there was a theory strongly anchored in linguistic.

In relation to **lexicon**, that is, the words that make up an indexing language, one should stress the need for their control, eliminating the synonymy, solving the homonymy, etc.

Regarding the **paradigmatic relations** – especially hierarchical relations recognized and established by different domains of knowledge – their greatest interest lies in the claim that they do not automatically represent the organization of a certain knowledge universe, but the goals of the information service, which reinforces the artificial character of the documentary languages.

However, since the 1960s, Gardin was concerned about the relations that can be established between terms of the lexicon, but do not constitute a paradigmatic relation, i.e., the syntagmatic relations, to the point of codifying them when developing SYNTOL (Syntagmatic Organization Language), as will be detailed below.

The optimization of the relation between original information (T1) and represented information (T2) can be analyzed through three unfoldments in Gardin's suggestions:

1. Proposing the construction of ontologies (section 1);
2. Adopting an extremely positivist view of science (section 2) and
3. In view of the cost involved in the proposals to be presented and discussed in sections 1 and 2, he imagined another more radical approach that proposed the reformulation of the scientific text, aiming to make it more searchable (section 3).

1. The proposal of ontologies in advance

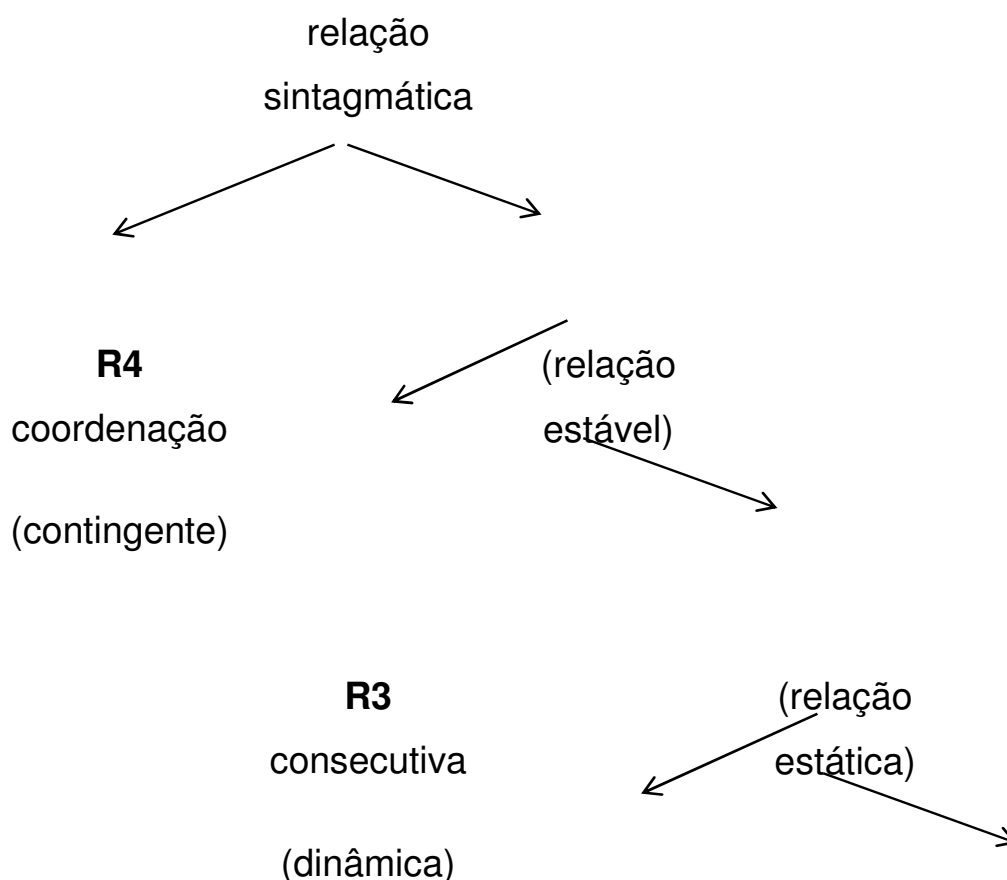
SYNTOL was developed in France by a group coordinated by Gardin, requested by EURATOM (European Atomic Energy Community) between 1960 and 1962.

SYNTOL is not an Indexing Language, but a general system of automatic documentation aiming the automatic indexing of texts, composed by an organized lexicon (that is, controlled and organized terms in a paradigmatic manner) and a quite

elaborated system of syntagmatic relations to establish the interconnection between the terms in the case of more complex issues. "Le SYNTOL est un ensemble de règles logico-linguistiques touchant différents manière les informations trouvées dans la documentation scientifique, en vue de leur exploitation sur calculateurs" (Cros, Gardin, Levy, 1964, p. 39-40).

Extremely detailed and very complex, the design of SYNTOL is very interesting from a theoretical point of view, but it is little used in practice. Tests were performed in 1962, for texts in physiology, psychology, sociology and ethnology and it was concluded that SYNTOL was "way out of the league".

If the practical implementation was compromised (both due to the complexity of the system and the technological limitations of the time) it was important to keep from the experience the precursor character of the proposed standardization of relations between the terms in the indexing timing. The syntagmatic relations were organized and encoded into logical sets (hence, the expression LSO – Logic Semantic Organization) and presented as different branches of relations. For example:



R2
Associativa

R1
Predicativa

The relation between “alcohol” and “joy”, for example, produced the syntagma < alcohol joy R3 >, where R3 denoted a consecutive relation. It should be noted that between the 1970s and 1990s, the use of *role operators*, or markers, was widely used. For example, working with texts from the medical area, the possibility of distinguishing a drug in the condition of a medicine to combat a disease or in the condition of provoking a disease was important. Today, it seems simple to associate what was proposed at that time with current ontologies; but its precursor character, with a still emerging technology, cannot be ignored.

The formalization of relations led to the definition of the minimal syntagma of any representation of information ($R_i(x, y)$), where R is a relation (paradigmatic or syntagmatic), x and y are terms of the lexicon. From the multiplication of ILs and from the development of multilingual thesauri, the notion of minimal syntagma showed its potential in studies of compatibility and interoperability between ILs.

One should keep from these proposals, dated back from the 1960s, the evident connection with the principles of artificial intelligence, also in broad development at that time, through the clarification of relations between terms and analysis of conditions in which relations can be established. The same theme is now updated in the development of ontologies, by identifying “objects”, properties, rules of relationships and applicability conditions: this set makes up the knowledge base and the inference engine (Gardin, 1989, p. 16-17).

2. Assessing information representation

The representational capacity of DA products received much attention from Gardin and certainly it represents a vein of research still little explored by TEMMA Group. The DA product assessment can be systematized from two complementary, but totally distinct, points of view:

- The assessment of the user “satisfaction” from a search in an information system. Precision, recall and noise calculations were highly developed by bibliography and of course they are very important;
- But Gardin considered another equally important calculation that was concerned with the representation level of the analysis product regarding the original texts (Gardin et al., 1981).

Both assessments should not be considered self-exclusive; but complementary although the second assessment, which emphasizes the representativeness of the representation in relation to the original text, is not common in our bibliography.

In my view, it overlaps the figure of the scientist to that of the documentalist. It is interesting to observe that the representativeness of text 2, produced from text 1 is, in practice, always presupposed, but not discussed. Gardin did not turn the issue into a moral or ethical question, but into a practical question: he considered that in order to effectively evaluate the result of their searches in the information system, the user of the information system should have access to all information related to the adopted DA procedures, as well as to all the details of the employed IL. Only the knowledge of all details that integrate the information process causes the user to leave a passive position – receiving the results of a search – to an active position of real evaluation of the results of the conducted search.

A user in a position to independently assess the result of a search and thus validate it: this is a challenge posed by Gardin that, in my opinion, seems far to be answered. But the challenge is beautiful, instigating and it deserves to integrate the research agenda. In this case, the user satisfaction is completed by information retrieval that corresponds not only to the terms used in his research, but also refer to texts that equally correspond to what was searched.

3. The logician project and the rewriting of the scientific text

In what seems to be the last phase of Gardin’s work, he asks himself more and more about the social meaning of scientific publications and information systems. In the 1980s and part of the 1990s, he invested a lot on the theme of text analyses, in different areas, trying to identify their consistency: at the end, he considered that the

texts on social and human sciences - filled with publications and libraries - presented many vices, but it was not possible to see advancement of knowledge. Obviously this conclusion is debatable, but it makes sense when placed into Gardin's perspective of total explanation of procedures for information representation. Not satisfied with this state of affairs, he elaborates new proposals, grounded in all his previous work, aiming at a more logical or useful solution for knowledge production in social and human sciences.

Approaching again the issue of documentation, without mentioning, he retakes the duality enunciated by Briet by identifying a primary order of texts, product of scientific and technical activity that is presented according to habits established by different knowledge areas and a secondary order in which the production of documentary information and creation of accesses to the primary order occur. Gardin considers that documentation has become hostage of the primary order, considering that the production of scientific texts has increased exponentially.

Then, he postulates that a "crisis of publications" is occurring due to a glaring imbalance between the production of scientific texts and the human capacity to consume them. The creation of tools to access the original documents – by improving the text selection procedures through indexation, elaboration of abstracts, alert systems, data mining and etc. - of course, this alleviates the imbalance between production and consumption, but does not solve the problem, because the capacity of information consumption by the human mind has changed very little over the centuries before the exponential growth of publicity of texts. "Malheureusement, cette quête d'une optimisation toujours renouvelée entre les silences et les bruits" remains even when it is softened (Gardin, 2001, p. 2).

The next question Gardin debates is more provocative: by assuming a totally adequate documentation system, which does not eliminate relevant documents in response to a query, combined with a well-balanced selection of relevant texts, one can postulate that all texts should be read since they are pertinent, but certainly only part of them will be fully read and many will be simply consulted. He mentions experiments carried out by specialists with text selection in a certain area to prove their relevance and time calculations necessary for the reading, concluding that it is impossible to read all the texts that are considered relevant. From this finding, Gardin proposes a remodeling of scientific texts in a way that they can be consulted considering that they will not be read. The "crisis of scientific publications" raises

another question: how many readers, on average, will have a scientific article? "est-il raisonnable de maintenir un système de publication d'articles dont les lecteurs sont si rares?" (Gardin, 2001, p.2)

From these questions, Gardin estimates that the problem is not in making the better texts available (documentation knows how to do this and does it very well), but in getting to consume them. One can, therefore, reshape Bush's statement when he proposed "the massive task of making more accessible our bewildering store of knowledge" (Bush, 1945, p.1) and state that it is no longer possible to continue writing texts to be read when most of them will only be consulted. While scientific texts are not remodeled, the documentation will continue as a hostage of the scientific publication system.

Therefore, it is an intervention proposal in the production of publications, decomposing them into different layers:

- Facts and data – material evidence, historical events, cultural practices;
- Theories, i.e., formulated views on facts or data;
- The adopted procedures to pass from facts to interpretations and theories.
- The cited examples to support the development of interpretations and theories.
- It is noticed, in this proposal, the return to concepts of knowledge base and of inference motor, which are cherished to specialized systems. According to his view (Gardin, 2001), the hypertext layers should:
 - Preserve the information content of the texts;
 - Explain the knowledge base and the rules to pass from T1 to T2 (rules of reasoning);
 - Exclude the methodological digressions, rhetorical resources, philosophical references, literary charm, etc.

The proposal was carried out by remodeling a 500-page book – on the pearls of carnelian, collected from archaeological sites in India and Near East and that lead to inferences about the societies that manufactured or imported them – in a multimedia CD-ROM containing the same informational content, with hypertext navigation, separating and relating facts, illustrations, supporting arguments, conclusions, schematizations of the adopted procedures in order to go from the facts to conclusions, etc.

The categorization of information types is surely a competence developed by the Information Science, which is why he proposes a cooperative work among information professionals, publishers, and authors; recognizing that the clarity about lexicons and relations between terms is characteristic of the information professionals.

4 – In summary

In addition to Gardin's contributions by introducing linguistics in classification studies, DA and indexing, we are left with the challenges that he launched:

- Recognize documentation as a linguistic stage of scientific work, by establishing terminology and discussing the relations among terms, according to fields of knowledge.
- Develop evaluation systems of the products in our work that combine the studies on satisfaction of our users and assessments on representation of our products in relation to represented texts.
- Invest on ontologies as a contribution to the industry of scientific texts publications in order to make them searchable. In this effort, the initial data and the inference rules must be fully disclosed: there is a lot of discussion about the availability of open data today, but the explanation of inference rules that suppose repeatability in the sciences is still a taboo in social and human sciences.

References

ALLARD, M., ELZIERE, M., GARDIN, J.-C., HOURS, F.(1963). Analyse conceptuelle du Coran sur cartes perforées. Paris: Mouton.

BUSH, V. (1945). As we may think. Atlantic Monthly, julho.

CROS, R.C., GARDIN, J.-C., LEVY, F. (1964). Le Syntol, um système général de documentation automatique. Paris: Gauthier Villars, 1964.

GARDIN, J.-C. (1989). Artificial intelligence and the future of semiotics: an archaeological perspective. *Semiotica*, vol.77, n.1/3, p.5-26.

GARDIN, J.-C. (1973). Document analysis and linguistic theory. *Journal of Documentation*, vol. 29, n.2, p.137-168.

GARDIN, J.-C. (1966). Eléments d'un modèle pour la description des lexiques documentaires. *Bulletin des Bibliothèques de France*, n.5, p.171-182.

GARDIN, J.C. (1974). Les analyses de discours. Neuchâtel: Delachaux et Niestlé.

GARDIN, J.-C. (1970). Procédures d'analyse sémantique dans les sciences humaines. In: POUILLON, J., MARANDA, P. Échanges et communications: mélanges offerts à Claude Lévi-Strauss à l'occasion de son 60ème anniversaire. Paris, Mouton, p.628-657.

GARDIN, J.-C. (1979). Une archéologie théorique. Paris: Hachette, 1979.

GARDIN, J.C. (2001). Vers un remodelage des publications savantes: ses rapports avec les sciences de l'information. In: CHAUDIRON, S., FLUHR, C. Filtrage et résumé automatique de l'information sur les réseaux: actes du Chapitre Français de l'ISKO. Nanterre: Université de Paris X, 2001, p. 3-8.

GARDIN, J.-C. et al. (1981). La logique du plausible: essais d'épistémologie pratique en sciences humaines. Paris: Editions de la Maison des Sciences de l'Homme.

Jean Claude Gardin and Document Analysis: trajectory of a semiology of representation



Maria de Fátima Gonçalves Moreira
Tálamo
University of São Paulo
mfgmtala@usp.br



Giovana Deliberali Maimone
University of São Paulo
gdmaimone@usp.br

1 Introduction

I do not remember when or where, I just know it was a long time ago. In an interview on TV, Fellini, responding to a question about how Rimini his hometown appears in his films, said he could not regard it as an objective fact: it is rather a dimension of my memory, when I am there I am assaulted by filed ghosts. Similarly, lecturing on the work of Gardin - the impact on the consolidation of studies on document analysis and on the transformation of the traditional view of operations in the field of information - implies in revisit it as affective memory, as a community of researchers created strong bonds of training, research and friendship around his work. It would not be, however, fair to equate it within these limits. Even to endorse the value assumed in the trajectory of researchers in Information Science, we decided to resize it, trying to revisit the theme on document analysis from its fundamental concepts, which confer an important reflective character to other appropriations that impact our view on the scientific work. For that, we rescued intervenient aspects on the proposition of the concept that still urges our perception and motivate intellectual curiosity.

That said, we intend to recover intervening aspects in the proposition of document analysis which relates, on one hand, to the characteristics of science and on the other hand, to the typological characteristics of semantic analysis. In other words, the approach on the concept will address the limits between two major universes: one related to the development of Human and Social Sciences until the 1970s, and the second strongly associated to the changes on document interpretation process, which, in Brazil, was object of fruitful studies, analyses and research that ended up shaping the formation of some professionals who are now invited to revisit some fragments of the emblematic text by Gardin "Procedures d'analyse sémantique dans les sciences humaines".

The concept of document analysis presents, from the meaning of its expressions - information management and systems, on the one hand, and museums on the other, a kind of conflict between past and present, between new and old, and at the same time the possibility of overcoming this conflict with the claim that these two times and worlds may converge around a single idea, relatively conciliatory and timely, but not always simple.

The trajectory of Gardin's thought, which culminates with the proposition of document analysis, has its origin in observing the impact of using more objective and

rigorous semantic analysis methods than the traditional instruments that guided the empirical interpretation in Humanities. In the 1970s, Gardin recognizes the importance of language in the production of science and the fundamental need for an interpretive methodological apparatus of linguistic nature.

With this, Gardin recognizes language as an area of scientific thought and not just a disciplinary area. It is common to all areas: for that, he cites the series of studies applied in various disciplines concerned with describing the significance of various texts. In short, Gardin shows us that we do not dispense language for investigative operations and how about rescuing the function of language?

To prevent a naive observation, it is highlighted that language does not appear as a result of thought, mere instrument of it, or as a means of communication. Gardin draws attention to the fact that language is representation. It is as a representation that fundamental methodological activities for investigative exercise is developed.

In this sense, every interpretation in the scientific universe participates or shares the representation function. In other words, all interpretations, theories concerning the world facts constitute a semiology. It is this concept that will be explored to understand the notion of document analysis.

2 Semiology of the scientific fact

In the Saussurean tradition, Semiology is defined as the science of signs. This, in turn, is traditionally presented, tautologically, as the relationship between a signifier - acoustic image - and a signified - conceptual image - which we call signification.

Although the definition of the term is difficult, some consensus are recognized: the sign is sensitive - the signifier - exists to a certain community of users and the signified - which, for the group to which the signifier becomes sensitive, points to a lack in itself. It is said, then, that the signifier may become sensitive and the signifier points to the absent part of the sign. Regarding the relationship between signified and signifier, named signification, cannot be confused with representation. Indeed, Ducrot and Todorov (1988) consider that the representation is associated with abstraction and this, in turn, can be stated according to different gradations or levels. It is inferred that along with the properties already mentioned, the sign system also presents representative properties. In verbal language, the sign is said, signifies, but, in addition, it also represents. The signification of the word "flame" does not prevent the representation of "passion/love" in the literature; and different degrees of abstraction

in Physics associated with expressions of source of heat and power supply, respectively.

It is observed in this example that the representation obtained by abstraction is embedded in the specific field of science and does not share the same arbitrariness of semiotic relationship in the common language. This means that scientific thought operates with a more abstract language - sign/representation relation, and therefore less arbitrary as the one between the sign and signification. The Stoics explained such associations through the distinction of three relations of the perceptible part of the sign: the speakable, the real thing and the mental image (DUCROT; TODOROV, 1988, p.105).

Probably this ambiguous or expanded aspect of the term sign is observed by Gardin, (1966, p. 171), when he defines it as a symbol

dans une acception délibérément générale, tout symbole désignant les éléments de l'ensemble lexical (mots-clé, descripteurs, termes d'indexation, etc.), que ces symboles soient empruntés ou non à une langue naturelle, « en clair » ou « codés », alphabétiques ou numériques, etc.

It is observed, in the transcription above, the use of the word symbol along with the strict definition of sign, that is, signifier relation - signified. This stems from resource widely used by structural linguistics to differentiate the strict definition of sign from other similar relations, but different that originate in their own canonical relationship.

This is the case of representation, which is associated with the symbol more than with the sign. Although used interchangeably, the two denominations serve different dynamics of word use associated not only to denotation and connotation processes, but also and mainly, to varying degrees of abstraction they undergo.

Thus, the relations between signified and signifier indicate union of different elements - necessarily the thing and the form: the sign is unmotivated. But the symbol has a peer relationship: usually both symbolizing and symbolized appear as peers of an unnecessary relation.

Here it is not the signification's universe, but the sense's. The universe of the scientific thought seems to be the symbolic universe, in which definitions and semantic analysis cooperate to develop the capacity of science representation. Gardin (19170) does not speak of the sign in the science universe, he asserts the existence of a

scientific semiology, guided possibly by semiology or semiotics, which proposes the symbol as representational unit of the system.

In the same way he clarifies the distinction between signification and representation, the arbitrary and motivated opposition applied to the sign does not dispense, for a proper understanding of the language of science and its expression in scientific texts, the distinction between fact and phenomenon (GARDIN, 1970).

The expression social fact refers to the object of sociology in its origin and serves as a resource for understanding the distinction between fact and phenomenon within scientific semiology. It is defined by a set of characteristics, it is built and cannot be reduced to any social phenomenon. When building its object, Sociology indicates a way of forming the disciplinary field of humanities and social sciences through an object of study, which we will name *fact* to harmonize with the general idea of scientific thought. Granger (1970) examines the conditions of existence of a fact, scientific/semiotic, as the product of a praxis (dynamic work) - and not as a passive subjectivity. From this perspective, reason develops a semiotic action, that is, a manipulation of signs.

It is in this perspective that Gardin (1970), by proposing a common reference - the scientific thought - for all activities that assign signification to the world, ends up associating it to a semiology. Thus, every interpretation is also a semiology, which makes us integrate semantic analysis to methodological apparatus of scientific practice from well-defined conditions of semantic analysis, listed as follows:

a) The object of analysis in this case is restricted to written documents, conceived as grammatical texts of a natural language. Similar empirical object can be subjected to empirical or paraphrased interpretation - understood as represented - by an autonomous meta-language;

b) The scientific thought as a symbolic system (semiology) designates a set of signs effectively built and enunciated and logically related.

The way to analyze scientific texts - which Gardin named comparative study - falls on the full set of signs. Specifically, this restriction is related to the precision of analytical operations, which must be explicitly presented and defined to tackle the passage from the original text to its representation.

The following results from those two conditions: that the success of the comparative analysis of semantic nature is linked to the identification of commonalities and differences between the system units, discriminated in particular symbolic systems. As the diversity of the semiotic universe is a fact, the analysis instrument must place the formal systems as methodological apparatus of analysis/representation of natural language and their products. Such discrimination is necessary since all science is structural, that is, turns the experience objective, possible to be presented as a network of interrelated elements. This way, it is ensured that the elements and the relations kept among them constitute an experience corresponding system.

However trivial it may seem, it is precisely the diversity of semiologies and its manifestation from the relation of signification that responds the crucial issue of semantic analysis: a word or a signification unit will never be univocal, which leads to the definition of *sign* by its representative function.

In this sense, it is worth mentioning the definition given to the expression semantic order. To Gardin (1966, p. 175) the expression comprises "un ordre des termes (c'est-à-dire des objets ou des notions qu'ils désignent) qui soit le reflet d'un ensemble de définitions couramment admises, dans un groupe humain donné." This reinforces the idea of scientific semiology and its symbolic units.

We also note that the symbolization process is not only about explaining the distinction between meaning and representation. Since the Saussurean structuralism, two forms of symbolic representation are distinguished and considered different: communication and symbolism. In Granger's (1970) perspective, the communication of a content is committed to living, signaling a concrete experience which makes reference - the perceptible part of the sign. In turn, the constructed form indicates a set of generic features, obtained by abstraction operations. From this perspective, both processes are irreducible but closely associated: the relationship between the interlocutors of a message is communicative, whereas symbolizing means the production of conceptual objects.

Therefore, the scientific object cannot be presented by the perceptible face of the phenomenon. This, although the result of a cut in reality, is descriptive - in it, its constitutive elements are isolated and defined. The semantic definition exemplifies this aspect since it presents the defining elements of the word through internal and positive properties. The most one may demand from the organization of a set of phenomena is their classification. A formal system, in turn, is characterized by relevant aspects of the

signs that compose it, unequivocally delimited and by a symbolic construction made from a closed list.

Thus, while natural languages exercise both communicative and symbolizing functions, in a balanced way, the formal systems have a higher level of abstraction and more limited communication, that is, restricted. In formal systems, the communicative function is secondary. It is understood, therefore, the meaning of terms such as "popularization of science" pointing to the "universal metalinguistic" character of natural languages, strongly associated with their primary model system function, and therefore, an efficient means of communication.

Briefly, Gardin (1970) recognizes that the scientific thought is developed through analysis procedures, also said interpretive. He also recognizes, with irony, that most times these interpretations are empirical. Empirical interpretations are associated with the totality of experience, i.e. they are related to what has been experienced, to the perceptible and do not account for the constructed fact. Of course, the semiology of living is the basis of all scientific practice, but is not enough to ground the science. The Semiology referred by Gardin demands the formalization of the living, or the recognition of its own logic, specific of the scientific practice, which should not be confused with the experience.

Thus, by proposing scientific practice as a Semiology, Gardin recognizes the need for a specific representation of contents of written texts, which should not be confused with the empirical interpretation that occurs most often, and which is submitted to subjectivity of meaning, of the fragments and the perception of experience. It is also noteworthy to remember that the consistency of the proposition of the notion of representation as apparatus of scientific thought - base of semiotics - does not privilege communication function, or rather, does not propose it as a primary function of the sign system.

3 Semantic Analysis and Document Analysis

Gardin (1970), as we have seen, parts from the notion of scientific thought to present semantic analysis as one of its methodological features. As mentioned, the dynamics of scientific thought, its universe, is a semiology. It is then necessary to clarify which procedures are therein registered.

Referring to the scientific thought in general, Gardin will not propose scientific semiologies associated with different areas of expertise. Although he does not flag the

discussion of interdisciplinary, his commitment is to identify the best methodological option as an adjunct of the scientific thought, or the linguistic operation able to define it. Thus, he recognizes a typology of textual analysis: content analysis, structural analysis (narrative) and literary analysis. He further notes that the formalization attempts fall on the first two, while the latter - literary analysis - was still surrounded by the tradition of textual criticism.

To contemplate a broader understanding of the role and procedures of the analysis in the construction of science, one must look at what happened in Europe, especially in France, the historicist and impressionistic spirit of the analyzes of humanity texts is compromised, from the 1960s, by structuralism in ethnology and linguistics. Levi-Strauss, Jakobson, Benveniste exploit, in their analysis of the cultural and literary fact, their narratives and textual structures that address their functioning (TRASK, 2004). In a certain way, the proposed analyzes of the facts of culture, understood as a whole, ultimately discredit empirical interpretations. As observed by Granger (1968, p. 138), in his study on myths, mythology is the object of analysis and not the myths. This is not a reflection on the world view of the people who created a set of myths. In this case what prevails is the subjective interpretation that is usually submitted to the appreciation by authority criteria: the analysis limits to the phenomenon and the fact is hardly formed or uttered. Mythology therefore includes the study/analysis of the fact and not of the phenomenon.

Of course, it is not just about validating or appreciating the quality of an interpretation or a comment made from data, texts or any other element of reality. It is rather about recognizing the arguments on which they are organized. Interestingly, the search for a pattern to guide semantic analysis is present in different stages of research on language. Almost invariably, the moves in this direction aimed at the study of literature, especially in the first decades of the twentieth century when Poetics starts to express scientific ambition. In fact, its object is not the set of literary works, but the literary discourse as a principle of engendering a multitude of texts (DUCROT; TODOROV, 1988, p. 84).

It is observed, therefore, that the empirical analyzes and interpretations have an important role in the development of Poetics, but do not constitute its core. Similarly, the notion of text - defined by its autonomy and closure - will eventually lead to different analyzes - narrative, thematic and rhetoric - each based on a specific approach to the text. In the case of narrative, for example, the analysis of folk tales performed by Propp

had fundamental importance to its definition as the referential text with represented temporality (DUCROT; TODOROV, 1988, p 268) which becomes the concept that guides the analysis itself, which addresses notions of transformation, mediation, action, character typologies and narratives, etc.

Although this context turns the duality between interpretation and analysis proposed by Gardin more evident, a more definitive explanation can be obtained from Granger. In 1968, in his work "Essai d'une philosophie du style," Granger develops a discussion on the process of structuring and ultimately defines the style as an individual solution caused by the difficulties encountered throughout the structuring work. This is because the transition from amorphous to structured is not done by imposing an external form: the scientific practice is the strength of the experience - of the living - and the structuring of experience can only be developed by work - the scientific practice itself - which relates form and content.

It also explains how the scientific work develops in mathematics and language, and how the individual behaves facing the structural solution: producing a style and not a subjective solution directly associated with the living. To cope with the distinction between the living and the structure - two different significant ways, he proposes two forms of sign manipulation, two semiologies, which we will call semiotics of living (of experience) and scientific semiology or structural forms. In Granger's terminology, semiotics of the living is a producer of sense and the scientific is a production of meaning. This opposition corresponds to what Gardin names, respectively, empirical interpretation and representation. The document analysis is a form of sign manipulation of representational nature, which constitutes the scientific semiology. This semiology, therefore, is informational.

A semiology of fact or of objective constitution of the significant system is, to Granger (1966), of informational nature as it treats reality informationally. The distinction established here is between what is structurable (*stricto- sensu*) and the significant systems. In fact, it is improper to admit that these two models can be developed within the semiology of science.

Considering what has been said, it is understood that the "interpretative analysis" of significant structures cannot be confused with the analysis of the symbolic structure (informational, of the fact). Indeed, the analysis carried out on the living (worldview) or significant analysis are not carried by the concept of structure.

Following this view, Garden (1970) proposes the document analysis as a manifestation of semantic analysis of the content of scientific texts "à la manière des documentalistes" (GARDIN, 1970, p. 630). For this proposal, he grounds on some techniques of automatic analysis of content or structure in anthropology. While traditionally the document analysis methods are reduced to empirical operations of assigning "key words" to texts, books, etc., founded more on experience and habit than the explicit procedure (GARDIN, 1970, p. 631), with automation, automatic analysis passes to identify the formalization of semantic interpretation. Thus, scientific texts are being substituted by paraphrases or by representations made by an autonomous meta-language, which defines the constituent elements of paraphrase.

Therefore, document analysis, through its objectifiable system, establishes a correspondence between the phenomena, or the living, and the meta-language representations. Document analysis then is a fundamental part of information Semiology, which in its development and transformation integrate the concept of knowledge associated with its production.

4 Conclusion

The perfect language, whatever the understanding, is part of the collective imagery and is often a pursued theme, probably because the Babel narrative makes us believe in the negative diversity - limiting and inoperable. As Eco (1993, p. 406-407) observes, discussing the consequences of language diversity qualifies it as an extremely positive phenomenon that allowed the establishment of settlements, the birth of nations and sense of national identity. Diversity is not inevitable and wishing to universalize it is an ambitious and innocuous process. Similarly, submitting a scientific paper to empirical interpretation is disqualify it as such, and impoverish it informationally.

The objectifiable semantic analysis recognizes underlying patterns of thought. Levi-Strauss (2003), one of the most important intellectuals of the twentieth century, by introducing modern anthropology, affirmed the need to consistently search for underlying patterns of thought in all forms of human activity. For him, men organizes knowledge in binary and opposite pairs that are structured according to the very logic of each universe, field, society, etc. This logic analysis shows, for example, that the

science text is based on the definitions of used terms. Such restriction of form, of logical nature, tends to facilitate formatting, storage and especially the use of the content of texts in a world where the researcher is increasingly overwhelmed by the volume of data. The representation in the sense given by Gardin, works with a control of the significance of living.

In other words, the scientist builds a text that responds to the requirements of argument transparency, covering solutions for different issues. The author's discourse depends undoubtedly on the forms of culture organization, but his scientific structure must be independent of the author's discourse, integrates the investigative discourse which is not, however, independent of the author. Under this view, the informational Semiology that develops logical analysis distinguishes two operations performed on the text: the interpretation of copyright nature, and the translation of scientific nature.

References

The references was made following the ABNT rules.

CABRÉ, M. T. *La terminología: teoría, metodología, aplicaciones*. Barcelona: Antártida/Empúries, 1993. 529p.

DUCROT, Oswald; TODOROV, Tzvetan. *Dicionário Enciclopédico das Ciências da Linguagem*. São Paulo: Editora Perspectiva, 1988.

ECO, Umberto. *A busca da língua perfeita*. Bauru: EDUSC, 2001.

GARDIN, Jean-Claude. Elements dun modele pour la description des leqxiques documentaires. *Bulletin des Bibliothèques de France*, v. 11, n. 5, p. 171 – 182, 1966.

GARDIN, Jean-Claude. Procedures d'analyse sémantique dans les sciences humaines. In: POUILLON, Jean; MARANDA, Pierre (coord). *Échanges et Communications*. Paris: Mouton, 1970.

GARDIN, Jean-Claude. Document analysis and linguistic theory. *Journal of Documentation*, v. 29, n. 2, p. 137 – 168, 1973.

GARDIN, Jean-Claude. Lecturas plurales y ciencias singulares de la literatura. *Diógenes*, Universidad Nacional Autónoma de México, 1982, v. 118, p. 5 – 16. Verano/1982.

GRANGER, Gilles-Gaston. *Essai d'une philosophie du style*. Paris: Librairie Armand Colin, 1968.

LÉVI-STRAUSS, Claude. *Antropologia Estrutural*. Rio de Janeiro: Ed.Tempo Brasileiro. 2003.

TRASK, R. L. *Dicionário de linguagem e linguística*. São Paulo: Contexto, 2004.

Documentary Language from J.-C. Gardin's viewpoint



Marilda Lopes Ginez de Lara
University of São Paulo
larama@usp.br

1 Introduction

Recent studies in Computer Science have been discussing 'adding semantics' to retrieval. This is not, however, a recent concern: Jean-Claude Gardin, about 50 years ago, sought to solve problems of a system or code whose terms should necessarily correspond to a sufficiently detailed definition to prevent doubts regarding its meaning (GARDIN, 1968). Gardin named this system 'meta-language', thus qualifying the language used to treat and retrieve technical and scientific texts.

Addressing issues of meaning remains a challenge, and revisiting Gardin's work may, not only disclose the theoretical and methodological basis of knowledge representation proposed or systematized by him, but also assess them in the face of current initiatives. It is also noteworthy to highlight the importance of his legacy for education: the Gardin's production makes a difference in academic and professional training in Information Science, as it ensures the understanding of the fundamentals of an activity which is often seen as merely instrumental.

In many ways, Gardin's reflections were ahead of his time presenting, in addition to deep knowledge of the literature on linguistic and logic, pioneering discussion and experimentation in the field of semantics. For today's research, Gardin is important because he addresses issues which have not been not fully solved, and because he conveniently places the problematic of meaning in the universe of representation and interpretation in language.

To organize the discussion, after this introduction, we will place the term 'documentary language' in comparison with other terminology and will discuss its classification as discursive metalanguage corpus to highlight the emphasis Gardin assigned to the relationship discursive metalanguage corpus. We then present a typology of lexicons, where Gardin compares document instruments to support lexicons, and then the main components of documentary language, characterized by Gardin as metalanguage. We try to disclose the author's discussion of semantic analysis, highlighting some of his comments on the relevance of the linguistics proposals at that time. In conclusion, we stress the relevance of his concerns, pointing to the relation of some of his proposals with those made recently.

2 Documentary Language: denomination and classification

The term documentary language divides its space with other terms of approximate meaning - indexing language, controlled vocabulary etc. - each one of them giving priority to a different aspect. More recently, the literature in English registers the term Knowledge Organization Systems - KOS, to group a number of resources and tools with different, but also close, functions: unstructured texts ('abstract' type), lists of terms and concepts (folksonomies, authority lists, tag lists, dictionaries, glossaries), structured lists of concepts (controlled vocabularies, classification systems, subject heading lists, taxonomies, thesauri, ontologies) and, finally, relationship structures (conceptual maps and reference models of organization forms) (SOUZA; TUDHOPE; ALMEIDA, 2012).

W3C opts for the generic term vocabulary, claiming that on the Semantic Web vocabularies define concepts and relationships or terms used to describe and represent an area of interest. Moreira Gonzalez (2011) also advocates the use of the expression 'semantic vocabularies' justifying the choice because the terminology is up to date in the Web context.

The variations due to different emphases, to the breadth expansion via KOS, or to specific update of the expression documentary language in this context, does not erase the basic difference that marks the use of the term 'language' in comparison with other denominations: the reference to natural language, complex communication system whose units have their value relationally established within the system, i.e., signify according to the set (Benveniste 1991). Vocabularies are not necessarily

structured: most often they are lists of occurrences of a given discursive corpus (GREIMAS; COURTÉS, 2008); on the other hand, documentary languages are, and their best characterization as such depends on the presence and sophistication of their semantic relations as well as the potential to generate syntagms via syntactical relationships. No wonder one of the most precise definitions of documentary language is the one defining it from its constituent elements: a lexicon, a paradigmatic network and a syntagmatic network (CROS; GARDIN; LEVY, 1968) - elements which are responsible for the structure of the instrument. The word 'documentary' (or documental, in some versions), restricts, in fact, the term 'documentary language' to a universe of documents, whose flexibility may be subject of discussion.

In the search for a model for Document Analysis, Gardin refers to three factors: "First, the definition of document analysis as the extraction of meaning from text implies a reference to something which is "taken out" of the text and designated by symbols which are not necessarily found in the text - namely, concepts and their interrelations, designated by *ad hoc* symbols (descriptors, role indicators, etc.) For the system of symbols thus used to express content, I proposed the name 'metalanguage' " (GARDIN, 1969, apud GARDIN, 1973, p.144). Second, the existence of similarity between existing metalanguages at the time, assigning their variations to differences between the fields and the representation objectives. Gardin identifies, however, relative analogies: a) the procedures followed to define the analysis unit, b) the establishment of analytical relations between the terms of the set, so as to construct the semantic organization of the lexicon in question, and c) the provision of syntactical relations between the terms to grant logical structure metalinguistic representation of the units under review. Third, the author refers to the distinction between analytic relationship (or semantics) versus syntactic (or logic).

The semantic relations are more stable - as the ones that occur in 'insecticides', unlike the 'unpredictable or syntactic' ones, which are circumstantial, accidental, such as in Drugs (instrument), Insects (Patient) (GARDIN, 1973).

However, the difference between the two has only practical purposes: under the syntactical plan, new relations are conveyed, which are object of speech and therefore not stabilized. In the field of science, the syntactical plan gradually passes to the semantic level: the new relations, in this sense, gain stability and then constitute semantic relations (GARDIN, 1973, p.144-145).

These factors together, justify for the author, characterizing documentary language as metalanguage, because it highlights the process of textual construction and transformation that keeps semantic contiguity with each other. This metalanguage has a normalizing function, operating semantically when treating homonyms, homotaxias, homographies, polysemies etc. (GARDIN, 1970). Although Gardin considers borrowing a word that has a different meaning in logic dangerous, the term metalanguage is interesting because it refers both to the starting language and to its representation (in language). These factors also justify his position in relation to the Chomskyan linguistic as the analysis unit and as to their contributions to provide the language with an articulation scheme.

A necessidade de uma metalinguagem, para o autor, se deve às 'anomalias' da linguagem natural do ponto de vista semântico: sinonímia, homonímia, homografia, polissemias; alotaxias (variações de formulação das frases), equivalências mais complexas (entre a designação e sua definição, por exemplo). Os sistemas simbólicos estranhos à linguagem natural apresentam irregularidade de correspondências entre significantes e significados fundadas na invariância relativa dos significados nos domínios considerados (GARDIN, 1970, p. 632). "La métalangage (...) n'est autre que ce système de symboles, aussi appelé 'langage documentaire' dans la littérature spécialisée ('information retrieval language' en anglais, ...). (GARDIN, 1970, p.634).

3 Opposition among lexicons: natural lexicography and document lexicography

Also important for the reflections on the knowledge and information organization is the proposal to position documentary languages in the set of 'lexicons'. This shows that they were not seen by Gardin as totally separated products from natural language. The typology, prior to the proposal of metalanguage, already shows the emphasis the author will give, in the future, to the relationship between natural language and documentary language and the specificity of the representation tools.

Documentary lexicons are understood as "tout ensemble de signes (mots naturels, symboles alpha-numériques, etc.), organisé ou non, servant à construire les représentations indexées de certains documents" (GARDIN, 1966). The products of 'document lexicography' and of 'natural lexicography' (glossaries and language thesauri), auxiliary of document activity, are now recognized as KOS.

In the construction of the trees that organize them, at each opposition, the differences between the considered lexicons are discriminated. The distinctive features that allow describing the dichotomies are the correspondence with the natural language (explicit or implicit), the organization or not of the lexicons (lists and classifications), the considered dimension (unidimensional and pluridimensional, real and apparent unidimensional), the type of privileged organization (semantic, syntactic or mixed). Under mixed, the classifications are organized by organization form (quasi-hierarchical, analytic-synthetic and faceted), and the first two organized according to the frequency of a term in the tool (univocal, multivocal).

Typologies are always interesting as a means to allow visualizations, but more importantly is that by using dichotomies in a growing process of semantic description, the structural nature of classification is reaffirmed, and its metalinguistic nature is anticipated.

4 Documentary Language Components

Focusing on documentary languages in detail, Gardin identifies in the lexicon, the minimum component of metalanguage, set of terms extracted from a given natural language (keywords or descriptors). A second component is the paradigmatic structure, set of data related to each other a priori in the lexicon (GARDIN, 1965 apud GARDIN, 1973). These two components characterized most documentary languages in the traditional configuration. A third element, a syntagmatic structure was added by Gardin and collaborators (CROS; GARDIN; LÉVY, 1968) in developing SYNTOL - Syntagmatic Organization Language, to provide greater flexibility and potential to the tool, as well as to cope with the immense diversity of related data observed in the process of document analysis (GARDIN, 1973).

The author relativized, however, the distinction between the two structures (paradigmatic and syntagmatic) stating that it was desirable to provide 'bridges' between one metalanguage and another when converting intra-metalanguage or inter-metalanguages (an anticipation of concern with compatibility between languages). The basic relational unit was then defined based on two structures, as the minimum syntagm 'R_i (x, y)', analysis algorithm for formalization, where x and y are terms of the lexicon and R is a binary relation whose meaning is specified by i. Guidance rules are added to avoid ambiguity.

Through this minimum syntagm, it is possible to represent both paradigmatic and syntagmatic relations: the former, stable relationships between terms; the latter, 'real' relations that presuppose interdependence between two meanings. Such relationships are classified as dynamic (consecutive or causality relations), static (associative and predicative superposition between concepts) or formal (coordinative, comparative and approximate).

SYNTHOL allows formulating sentences by connecting pairs of words based on their mutual relations of dependence. The system predicts the indication of the type of existing logical relations between considered words (predicative, associative, consecutive or coordinative relations), signaling the direction of reading the syntagm and determining their formal categories (determining the category in a thematic level - place, time, manner - wider than the level of syntagms in order to make the context clear). Syntactical operators are used to avoid interpretation ambiguity, when necessary, or to connect nearby syntagms (CROS; GARDIN; Lévy, 1968).

The proposal set out, among other observations, from the finding that the binary relations are often used by metalanguages to perform semantic organization in the same way that role indicators do for expressing syntactic relations: the functional markers (unary predicates), each one implicitly referred to one or several semantic markers in a given metalinguistic representation. According to Gardin, the advantage of the binary format is that it can guide logical calculations to make deductions and inferences in the search process when lexical or syntactic specifications should be gradually expanded to increase the recall in retrieval or, in converting one metalanguage into another, when the mappings (syntagmatic trees or SYNTOL graphs) of the two imply different relation trees (GARDIN, 1973).

The SYNTOL experience, presented here, very briefly shows the progress of research in Knowledge Organization and Representation at a time when most syntagmatic operations were, at best, resolved with the use of Boolean logic.

5 Semantic Analysis

By searching for procedures to evidence the meaning of various texts, Gardin (1970) shows, already in the 1960s, the importance of semantic analysis, when the methods of document analysis were reduced to empirical assignment of keywords assignments to a text. The author discusses the limits of Chomskyan linguistics - the

transformational generative theory - and complains that linguists rarely speaks of semantic analysis.

Gardin criticizes that the dominant linguistic theory at the time gave priority to syntactical analysis and to the 'surface structure', which would exercise a priority role in determining the meaning. He talks about the limits of deep structure, questions the sentence as a standard unit of analysis, the use of grammatical categories (nouns, adverbs, adjectives, etc.) and of grammatical functions (subject, object, etc.) as a means to express meaning. He cites other authors (whose contributions were published between 1968 and 1972), who also questioned the Chomskyan proposals, such as Wilks and Hutchins.

On the other hand, Gardin refers to a series of experiments conducted by linguists at the same period (beginning of 1970s): the pre-lexical structure proposal (Post), the adoption of 'semantic strategies' to guide the interpretation of sentences (Bever), the definition of 'formational structures' prior to statements analysis (Grimes), a smaller unit of analysis than the phrase (Wilks) and, instead, larger units to provide a better interpretation (suggested individually by Langendoen, Lakoff, and Thompson). The author highlights the research developed in the former USSR - Mel'chuck and Zholkovski, among others - who proposed sophisticated language models based on semantic structures and who, unlike Chomsky working from text to meaning, went in an opposition direction, from conceptual frameworks to linguistic expressions (GARDIN, 1973).

Gardin also reports other experiences he found interesting, such as discourse analysis, the appreciation of the 'logical proposition', logical concepts as concepts that underlie the speech, the semantic networks that incorporate relational or combinatorial information 'on the world' (Lakoff; McCawley, Kuroda). Similarly, he refers to 'assumptions' of language, or selection rules that guide the linguistic behavior (citing Bever, Fillmore and Langendoen), which the author places in parallel to the notion of semantic or paradigmatic organization used in document analysis. In addition, he also refers to the categories used in formulating prepositions (originally the case grammar proposed by Fillmore in 1968) reformulated by Langendoen (with 'roles structures'), by Bever ('basis relational structures'), and by Wilks ('mental templates').

All the criticism made to the transformational generative linguistics was in favor of logic, not the philosophical logic, but to symbolic, technical logic. It would provide the appropriate means to express the organization of thought, to categorize

vocabularies (Agent, Patient, Instrument, Objective, etc). Regarding categories, Gardin highlights the documentary practice as anticipatory of many of the proposals for an applied semantics.

6 Conclusion

As we observed, Gardin contributes to the understanding and development of documentary languages and to the representation of technical and scientific texts. His work, however, is not accepted without criticism: while he was a man ahead of his time - realizing in semantics the problems for conveying meaning, emphasizing the importance of structuring metalanguages etc., - we should remember that he was inflexible as to the approach on human sciences discourse, criticizing them for their lack of objectivity and clarity from the organization reference of natural sciences discourse. Gardin saw the expression of richness and complexity of the vocabulary in human sciences discourses as 'abnormalities', looking for means to achieve ways to prevent the rhetoric excess that characterized it.

Undeniable, however, is the author's contribution to the improvement of the procedures with documents as if it were necessary to experience a rigorous phase to review it again, afterwards (as Wittgenstein's statement, in the *Tractatus*: "He must so speak throw away the ladder after he has climbed up on it" (WITTGENSTEIN, apud GRAYLING, 1966, p.65).

Finally, I highlight the actuality of Gardin's project when observing the similarity of principles underlying the proposal of the syntagm 'Ri (x, y)', made in the 1960s, and the RDF - Resource Description Language, language for representing metadata in the form of sentences that relate entities and properties. Not least interesting are his concerns about metalanguages compatibility and mapping, which now correspond to initiatives to promote interoperability between vocabularies.

Many other of his contributions were left out of this presentation, but not less important and related, which followed the reflections on metalanguages and experiences about automation of analysis. One of them is the proposal for schemes of reasoning or logical links between descriptive propositions (facts and arguments) and interpretative propositions. Very briefly, it is a system that, from a knowledge base, goes from data to interpretation through a logical apparatus of calculus, an inference engine, as an expert artificial intelligence system (GARDIN, 1987). Product from logician analysis, it is a device of greater formalization and sophistication that allows

to pass from a simpler indexing to representation and to inference (and interpretation).
But that is another subject.

References

The references was made following the ABNT rules.

BENVENISTE, É.. *Problemas de linguística geral, I*. Campinas: Pontes; Ed. UNICAMP, 1991.

CROS, R.-C.; GARDIN, J.-C.; LÉVY, F. *L'automatisation des recherches documentaires: un modele général "Le SYNTOL"*. 2. ed. revue et augmentée. Paris: Gauthier-Villars, 1968.

GARDIN, J.-C. Elements d'un modèle pour la description des lexiques documentaires. *Bulletin des Bibliothèques de France*, 1966, p.171-182.

GARDIN, J.-C. Procédures d'analyse sémantique dans les sciences humaines. In: POUILLON, J. et MARANDA, P., orgs. *Échanges et communications: mélanges offerts à Claude Lévi-Strauss à l'occasion de son 60ème. anniversaire* (Tirage à part). Paris: Mouton, 1970, p.628-657.

GARDIN, J.-C. Document analysis and linguistic theory. *The Journal of Documentation*, v. 29, n.2, p.137-168, June 1973.

GARDIN, J.-C. Éléments d'un modèle pour la description des lexiques documentaires. *Bulletin des Bibliothèques de France*, v.11, n.5, p.171-182, 1966. Disponível em: <<http://bbf.enssib.fr/consulter/bbf-1966-05-0171-001>> Acesso em jun.2015.

GARDIN, J.-C. *Systèmes experts e publications savantes*. London: The British Library, 1987.

GRAYLING, A.C. *Wittgenstein*. São Paulo: Ed. Loyola, 1966.

GREIMAS, A.J. & COURTÉS, J. *Dicionário de semiótica*. São Paulo: Contexto, 2008.

MOREIRO GONZÁLEZ, J.A. *Linguagens documentárias e vocabulários semânticos para a web: elementos conceituais*. Salvador: EDUFBA, 2011. 128p.

SOUZA, R.R.; TUDHOPE, D. ; ALMEIDA, M.B. Towards a taxonomy of KOS: dimensions for classifying Knowledge Organization Systems. *Knowledge Organization*, v.39, n.3, p.179-192, 2012.

From Document Analysis to terminology: theoretical and methodological trajectory



Vânia Mara Alves Lima
University of São Paulo
vamal@usp.br

The aim of this paper is to present the theoretical and methodological trajectory on the use of knowledge domain terminology as a reference to the construction and management of documentary languages, specifically controlled vocabularies and thesauri.

This research problem begins to emerge in the 1990s during our professional practice in the technical processing of one of the largest library collections of São Paulo University, (USP) the Library and Documentation Department of Philosophy, Languages and Human Sciences, which provided us with daily issues about information organization, representation and retrieval, produced on the same subject in different areas of knowledge, and which consequently addressed different views.

When we were transferred to the Library and Information Service of Faculty of Architecture and Urbanism of USP, we faced another problem related to information representation, the knowledge produced in the same domain was represented by different designations, often depending on the support on which it was registered, i.e., books, maps, architectural designs, slides, photos, journal articles.

The search for theoretical foundation to solve practical problem led us to the works of Smit, Tálamo, Cunha, Lara, Kobashi, Guimarães and Fujita, founders of Temma Group, which have the work of Jean Claude Gardin as theoretical foundations, but which considerably expanded the horizons of the so-called document analysis, its processes and its products, especially as regards the construction of documentary languages.

The studies by Temma Group showed that Document Analysis defined by Gardin (1981) as "L'expression désigne, on le sait, un ensemble de procédures pour exprimer le contenu des documents scientifiques sous des formes destinées à en faciliter le dépistage ou la consultation" is a methodological discipline that suggests procedures for text analysis in order to select information contents which may be represented, retrieved and disseminated (TÁLAMO, LARA, Kobashi, LIMA, 1992, LARA, 2011) regardless of the support on which they are recorded. To the concept of documentary language as "Un ensemble de termes utilisés pour représenter certains contenus de documents scientifiques avec des fins de classification ou de l'information de recherche rétrospective" (CROS, GARDIN, LEVY 1968)", Temma Group adds other characteristics that emphasize its character of "language" as a structure, whose terms must necessarily be related so that they may mean, in a certain way, working as a communication vehicle that represents the conceptual domains, respecting the community culture which it serves (VOGEL, 2004).

In this framework, we address our linguistic and terminological approach both on the development of documentary information, understood as the result of content representation of a document and the use of domain terminology as a reference for the construction of documentary languages.

Regarding documentary information, we try to outline it from the approaches of linguistic references, identifying it as a documentary sign, similar to linguistic sign, as defined by Peirce (1977,) as something that, in one sense or way, represents something to someone and is in place of something else in some respect or capacity, i.e., documentary information is proposed in place of recorded knowledge and, therefore, in the same way the linguistic sign is capable of semiotic process denominated by Lara (2006) as documentary semiosis.

Understanding semiosis as the construction of meaning by the interpreter, within a given context, it can be inferred that documentary semiosis is the construction of meanings based on terminological references which refer to the conceptual structures of domains. Therefore, contextual references of documentary information production are essential so that they effectively represent a set of true statements about the recorded knowledge, which is compiled in the definitions of each term present in the terminology of a domain.

We must clarify that when we talk about terminologies we are referring to terminology as a product, that is, the set of terms of a specialty, which is developed

through the use of terminological standards proposed by Terminology as a discipline that addresses specialized terms (CABRÉ, 1995).

While domain terminology is the formal reflection of the conceptual organization of a specialty, and inevitable means of expression and professional communication which ensures the transmission of knowledge, documentary language has the function to normalize the search and ensure the retrieval of this recorded knowledge by preparing documentary information. Thus, it is understood that these two instruments are complementary.

According to Gardin (1981), documentary information is the product of document analysis, the result of a semantic operation formulated within a documentary language that transforms an original text in one or more keywords, and that even presenting - in the documentary language - the same form in natural language, does not necessarily have all the meanings present in a general language dictionary.

In this context, we infer documentary information as the content representation of a document, from the domain concepts to which it belongs, which designated by the terms of this domain serve as a reference for the descriptors of documentary language because they contextualize its meaning from a practice in this domain (LIMA, 1998).

In the following figure, we depict documentary information represented by the triad concept/term/descriptor, result of semiosis documentary process coupled with the practice in a knowledge domain.

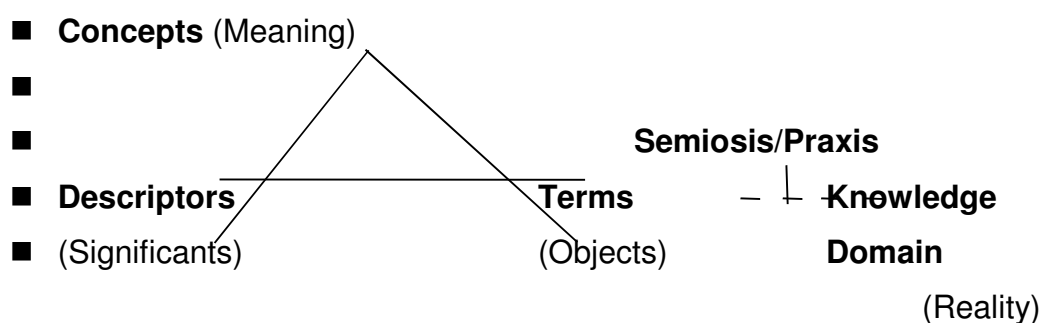


Figure 1 – Documentary information

This scheme of documentary information meets Smit's (2000, p.28) arguments that, for documentary information to work effectively and correctly as a mediator between the user and information stock, the user must be able to contextualize it and decode it, because when using the domain terminology as reference for documentary

language, the user may recognize the terms that are part of his/her practice. In other words, coding documentary information from references that will be recognized, decoded and interpreted by the user in semiosis documentary process.

Guarantee of quality in information representation and retrieval depends on the effective coding and decoding of documentary information, but it is not limited to understanding its structure, it should also consider some dialectical tensions between existing opposing forces when management and maintenance of documentary language, such as **conservation x mutation** and **consensus x specificity** (LIMA, 2004).

Conservation ensures the understanding among subjects, and **mutation** meets the changing needs of society, i.e., in the management and maintenance process of documentary language, one should maintain access points from previous systems and at the same time that should enable the addition of new access points (LIMA, 2004).

Documentary language must also meet both **consensus** and **specificity**, that is, on the one hand, it should meet its characteristic of being institutional, as it is always built to be used under the objectives of a particular institution ensuring mutual understanding of the subjects, and on the other hand, it should provide elements of a specific experience to each user (LIMA, 2004).

The need for documentary language to ensure consensus at the same time that meeting specificity demands precision and consistency from descriptors that can only be achieved from the compilation of true statements that will disclose their meaning and disclose semantic relationships that articulate their conceptual network. In other words, it is necessary to identify the characteristics that make up the concept, whose designation (term) serves as reference for the descriptor who constitute documentary information when assigned to a document. Finally, the sense of documentary information is expressed by the definition of the concept it denominates.

In turn, the definition of the concept gathers attributes or characteristics that allow to determine the categories of a documentary language. These categories are defined by the common trait of a whole class of concepts/terms/descriptors which, for this reason, are associated. The limit of each category is established by specific traits that allow to individualize each concept/term/descriptor establishing the disjuncture among the elaborated documentary information.

Here, we include in the trajectory the notion of semantic class that was also addressed by Gardin (1966) when describing documentary lexicon referring to the need to organize its terms to disclose the existing hierarchical relationships among them, either by affinity or by semantic difference.

Determining the semantic domain in linguistics is, according to the epistemological assumptions, seeking to discover the structure of a given domain, or propose a structure to it (Dubois et al, 2011). Trier, cited by Lopes (1987), observes that the lexical units of a language are organized into structured groups in such a way that each unit is defined therein by the position it occupies in relation to the others. Thus, we infer that the meaning of a concept/term/descriptor is specified by its similarity and its difference in relation to other relevant elements of the semantic domain, as one word only acquires its meaning as in opposition to other units in the same field (Germain, 1981).

Genouvrier and Peytard (1974), with regard to the semantic domain, defined it as the set of employments of a word (or syntagm or lexia) where and through which the word acquires a specific semantic load and the delimitation of these employments would occur by recognizing all immediate contexts that the word receives in a given text.

According to Hernando Cuadrado (1995) the minimum condition for the words to belong to the same domain is that they have a significant common trait (sema) (the higher the number of semas, the more coherent the semantic field will be, and in general, the fewer words integrates it); a word can take part of all semantic domains that are built over any significant traits that are discovered in it; when a word has several meanings, each of them belongs to a different semantic domain.

For example, while we can identify as belonging to the semantic domain of the word *table*, due to the common characteristic "object that allows gathering around", the following words: dining table, round table, assembly table, operating table, each may form part of a different category in a documentary language, due to delimiting characteristics such as "for foods"; "for discussion"; "for surgery", etc. In another case, the word *iodine* may to be included in different categories of documentary language from delimiting characteristics, namely, *to be a raw material*; *to be a product*, *to be a reactant*.

At the moment one has discussed interoperability among documentary languages in semantic web, we believe it is necessary to deepen discussions on

mapping of semantic fields that make up a domain, as only from concept characteristics listed in the definitions of terms, that will serve as referent and will contextualize the descriptors of documentary languages, it will be possible to make information representation and retrieval more effective.

References

The references was made following the ABNT rules.

CABRÉ, M. T. La terminologia hoy: concepciones, tendencias y aplicaciones. *Ciência da Informação*, Brasília, v.24, n.3, p.289-298, set./dez. 1995.

CROS, R.C.; GARDIN, J-C; LEVY, F. *L'automatisation des recherches documentaires: un modele general: le Syntol*. Paris: Gauthiers-Villars, 1968.

DUBOIS, J. ET AL. *Dicionário de linguística*. São Paulo: Cultrix, 2011.

GARDIN, J. C. Éléments d'un modele pour La description des lexiques documentaires. *Bulletindes Bibliothèques de France (en ligne)*. n.5, 1966. Disponível em <http://bbf.enssib.fr/consulter/bbf-1966-05-071-001>, Acessado em 20 jun. 2015.

GARDIN, J. C. *La logique du plausible*. Paris: Editions de la Maison des Sciences de l'homme, 1981.

GENOUVRIER, E., & PEYTARD, J. *Linguística e ensino do português*. Coimbra: Almedina, 1974.

GERMAIN, C. (1981). *La sémantique fonctionnelle*. Paris: PUF, 1981.

Hernando Cuadrado, L. A. *Introducción a la teoria y estructura del lenguaje*. Madrid : Editorial Verbum, 1995.

LARA, M. L. G. Conceitos de organização e representação do conhecimento na ótica das reflexões do Grupo Tema. *Inf. Inf.*, Londrina, v. 16. n. 3. p. 92 – 121, jan./ jun. 2011.

LARA, M. L. G. É possível falar em signo e semiose documentária? *Enc. Bibli: R. Eletr. Bibliotecon. Ci. Inf.*, Florianópolis, 2º número esp., 2º sem 2006

LARA, M. L. G. *Representação e linguagens documentárias: bases teórico-metodológicas*. (doutorado), Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo, 1999.

LIMA, V.M.A. *Da classificação do conhecimento científico aos sistemas de recuperação de informação: enunciação de codificação e enunciação de decodificação da informação documentária*. (doutorado), Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo

LIMA, V.M.A. *Terminologia, comunicação e representação documentária*. 1998. Dissertação (Mestrado) Escola de Comunicações e Artes, Universidade de São Paulo, 1998

LOPES, E. *Fundamentos da linguística contemporânea*. São Paulo: Cultrix, 1987.

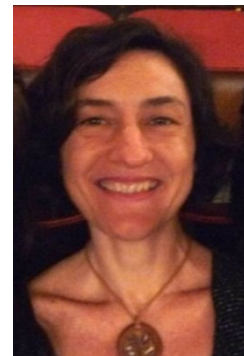
SMIT, J. W. Informação. In: LIMA, Y.D.; SMIT, J.W. (Org.) *Organização de arquivos*. São Paulo: IEB/ECA, 2000. p.19-31

PEIRCE, C.S. *Semiótica*. São Paulo: Perspectiva, 1977.

TÁLAMO, M.F.G.M.; LARA, M.L.G.; KOBASHI, N.Y. & LIMA, V.M.A. Instrumentos de controle terminológico: limites e funções. *Actas do II Simpósio Latino-americano de Terminologia*, 1990. Disponível em: <<http://www.ritem.net/actes/2simposio/indice90.htm>>. Acesso em 20 de junho de 2015.

VOGEL, M.J.M. A influência de Jean-Claude Gardin e a linha francesa na evolução do conceito de linguagem documentária. *Perspectivas em Ciência da Informação*, v.14 , número especial, p.80-92, 2009.

Ordering Documents: underpinnings and relations with bibliographic classification



Cristina Dotta Ortega
Federal University of Minas Gerais
ortega@eci.ufmg.br

1 Introduction

Ordering documents is the activity that allows us to propose an arrangement for a collection, relevant to the institutional objectives, the characteristics of documents and activities of the target audience. This arrangement has the objective to provide a way of reading the collection that is made possible by the circulation around the space, to grant access to documents through its location from codes listed in database records, and documents arranged in the collection, and to enable the management of a collection with respect to selection policies, space optimization and preservation of documents. These functions can be operationalized by an ordering plan, an instrument that provides a structured map to the spatial arrangement of the collection, including the codes to be used for the documents.

An activity performed in archival, museological and bibliographic approaches, document ordering will be treated here only regarding the bibliographic approach. Although difficult to define, it seems to us, we can say that the bibliographic approach of information includes: information for knowledge construction, necessary to educational, scientific and professional activities in general; information for enjoyment or aesthetic experience; and information for useful purposes, on access to services or entertainment activities, education, culture, health and civil rights in general. The bibliographic perspective also involves sensory experience from objects in exhibitions

or similar activities, which are not discussed here, emphasizing the cognitive character of the messages produced on objects.

The problems justifying the theme of this paper are described below.

Despite the need to adapt procedures to their contexts aimed at communicating with the public, one way to perform the activity of ordering documents prevails over others. This is the call number, created with the purpose of a document location code in a collection. The call number belongs to the set of technical processing solutions of library collections that make up what we call the Anglo-American model. Extensive adoption of the call number by the library community is carried out by use of the Dewey Decimal Classification (DDC) or the Universal Decimal Classification (UDC) for classificatory notation, whose function is to group documents, followed by the number corresponding to the name of the author, usually the main access point of the card catalog according to the Anglo-American Cataloguing Rules (AACR2), using the Cutter table of author names, with the function of individualizing the document. The call number model gained greater legitimacy and dissemination capacity in Brazil by being part of the syllabus in undergraduate courses in Library Science. However, generally, historical, theoretical and methodological aspects underlying the call number are not examined in-depth in educational programs. Moreover, an ideal program would begin with the concept of document location code, thus capable of adapting to different contexts.

The unique model gives precedence to the use of the cited traditional bibliographic classification systems, the DDC and the UDC. Regarding the instruments as components of a model, several points are worth deepening. First, the issue that bibliographic classification systems do not allow ordering, as they do not perform alone the individualization of the document in a collection. Furthermore, ordering is not only performed by the classification method, it may occur by sequential method, or both. Thus, an operation is taken by another, namely the classification takes the place of ordering. The second point is that bibliographic classification systems are studied from the perspective of thematic representation of documents, together with the lists of subject headings and thesauri. However, the attributes of documents relevant to a physical arrangement, which are clearly indicated in the traditional bibliographic classification systems and in the way they are used, do not refer only to subjects. A third issue is the understanding that the thesauri would be more relevant for information retrieval, compared to a lower weight that bibliographic classification systems would

perform as document orderers. As an activity does not overlap the specificity of each other, ordering continued necessary, but both the process itself and the instruments adopted to it have not been sufficiently studied. As a fourth and final question, we point out the problem of assimilation of the concept information organization by classification. Although, in a broad sense, all procedures of information organization involve classificatory cognitive operations, they are not restricted to these operations.

As stated previously, we verified the existence of gaps in Information Organization on processes and classificatory ordering instruments, but not limited to them, which hamper its development as the understanding of a process occurs to the extent it is fully explicated, as well as its relationship with other processes.

Thus, the aim of this study is to explore the foundations of bibliographic ordering of documents through the study of the specificity of the process, their instruments and their relations with the process of bibliographic classification. The methodology included reading and selection of texts on underpinning concepts of document ordering process and its implementation, integrated to other processes. Texts in Portuguese (from Brazil and Portugal), Spanish, English and French were selected and adopted for research. We adopted, as the main reference, the contemporary French literature on the subject, developed around the term ordering (*classement*, in its original form in French), for its greater generalization about the processes, instruments and products that make up the activity, but also for being in use in French institutions. About the methods of ordering - sequential and classificatory - a brief systematization is proposed from scattered indications in the literature, basically Salvan (1962), Jouguelet (1997), Maniez (2002), Carrión Gútiéz (2002), Calenge (2010) e Combout (2011). The study of classificatory instruments of ordering and of traditional systems of bibliographic classification was based especially on Gardin and Coyaud that, from their own texts, both published in 1966 in France, addressed the linguistic approach as underpinnings for classificatory schemes, from the idea of unidimensionality that is intrinsic to them. Based largely on this approach in the context of studies performed by the Temma Group, from Escola de Comunicações e Artes (ECA) of USP, Lara (1999, 2001 and 2002) discusses the linguistic aspects of traditional systems of bibliographic classification, in addition to specifically observing their pragmatic aspects.

About the structure of the paper, in topic 1 we contextualize document ordering as an information organization process. In topic 2, we develop the conceptualization of the document ordering process, through their related processes, methods,

instruments and products. In topic 3, we discuss the specificity of bibliographic classification in order to advance understanding of document ordering, noting the uniqueness of each one of them and the interconnections between them. In topic 4, we carry on about the linguistic foundations of classificatory instruments of ordering, analyzing the best known bibliographic classification systems. We present, at the end, some conclusions on the subject.

2 Contextualizing document ordering

As a documentary activity, ordering documents is carried out in the context of information mediation actions. The appropriation, which materializes through these mediation actions, must be addressed, according to Perrotti and Pieruccini (2007, p. 60, 88), within a broad framework of questions involving information and its facilities (libraries, centers and documentation, information, and memory institutions, among others), in its multiple interactions and angles. We talk about mediation between objects and people related to the object confrontation, according to Meneses (1994, p. 30), leading to its transformation into a document, and the observation of people as potential users of information. Mediation is a symbolic process that operates on objects and people in the perspective of documents and users, so both documents and users are defined situationally.

The people we name users are those addressed in some dimension of their lives from which certain information can be used with advantage, be it scientific, useful, or other. Although all people are users of information, since the need and use of information are inherent to human beings, the area works with these particular information needs, aiming to satisfy them, but also raising awareness. So it is about audiences, more than users, meaning people who are recipients of the production of messages about objects.

Documents are objects addressed informationally. The document is defined by a physical instance (material medium) and a symbolic instance (informative, content related), the latter always considered as an assignment, meaning it is not inherent to the object. Thus, the informative dimension of the document is not previous, it does not exist *a priori* or *in natura*. It is the defining instance of the existing condition of the document, it is constructed at the time of interpretation or 'making itself a document', even though the physical dimension is required for this. For Meyriat (1981, p. 54) and Sagredo Fernandez and Izquierdo Arroyo (1983, p. 265), 'to be a document' depends

on being used as such. Meyriat (1981, p. 51) also states that the document can be defined as an object of information support which serves to communicate, and which is durable, so that communication can be repeated.

If the document is a constructed object, the bibliographic document, which we address in this text, is not a pre-existing object, but it depends on human actions of meaning that make it as such. We must overcome the belief that the bibliographic document is pre-defined by its typology, for instance the book and the journal, these types are often in bibliographic information systems such as libraries, documentation centers and specialized databases, among others, for including scientific or aesthetic content. Another review that is urgent, is the idea that the medium would change the document type; on the contrary, a book, by presenting a certain type of message conditioned by a form of content organization which characterizes it, is still a book even if it is produced on paper or electronically.

Since no object is itself a document, the transformation process of the first into the second is a result of documentary activities that carry out the production of messages about objects in users' view. Thus, from the point of view of the area that concerns us, an object is made document from these activities.

Another moment, however, is the one about the interpretation made by users: this interpretation is influenced by the production of messages about objects, but not determined by it. There is no direct correspondence relationship between messages about objects and the interpretation of them, but the observation of the various appropriations carried out by the users (or not) contributes as a way to assess and review the documentary activities.

Through documentary activities, documentary information systems or documentary systems are produced. Among them, we cope with bibliographic information systems. Initially characterized by bibliographies and libraries with their collections and catalogs, now they have a greater range of variations depending on the typologies and documentary mediums, the contents of documents and audiences and activities in which they inscribe.

Therefore, we have several bibliographic repertoires as bibliographies, library catalogs, databases, information portals and hierarchical structures for navigation and online resource access, among others.

In general terms, we can consider as documentary activities: identifying information needs; selection; collection; representation; ordering; preservation; and

development of information products and services, exhibitions and educational and cultural activities.

To identify information needs, a diagnosis of the production flow and use of information in the context of activities is performed. From the diagnosis, it is possible to obtain criteria that support the selection of documents.

The product of the document selection is the collection of documents, which is determinant in setting up the system for providing its personality. The documents of the 'collection' are collected or not. Documents cannot be collected when there is an indication of their electronic address or the name and location of the person or the same depository institution; there are also databases that do not have location information for access, as in the case of scientific databases, especially in the period before the publication of journals in electronic media. When documents are to be listed, their ordering is performed through codes, or not; when performed, the ordering of documents is the last stage of information organization, the one in which the documents become finalized for access and use by the user.

The document representation occurs in the documents themselves - in the case of codes assigned to them for their ordination in a certain space - and/or bibliographic repertoires, such as databases, in the form of records - in which the ordering of its access points occurs.

The activities for document preservation are required to provide possibilities for further uses.

In general, depending on the previous activities, information services and other actions and related products are made, such as: systemic and contextual search of information in different systems or in none of them; access to the document (except registration databases, in which the record must answer the query); selective dissemination of information; support for information use and training of users; the production of exhibitions; and various educational and cultural activities for public awareness.

Among the documentary activities mentioned, those that specifically make up information organization are described below:

- document selection;
- document collection;
- document representation in the form of component records of bibliographic repertoires, through the processes of description, indexing and summary preparation,

and the choice and form of access points, and in the form of document location codes, often using bibliographic classification process; and

- document ordering and document metadata ordering from codes of documents, in the first case, and access points to various bibliographic records repertoires, in the second.

As we can see, ordering is a broader process involving more than the activity of physical allocation of documents. Ordering was defined, in a broad sense, as early as 1962, but most likely before that, as an activity involving documents, records from a catalog or references of a bibliography or a print catalog (SALVAN, 1962, p. 1). Thus, in all the procedures that make up the organization of information, ordering must be understood from its two facets: the ordering of documents and the ordering of document metadata. In the first case, there is the physical layout of documents. The second case mentions metadata, term currently used to indicate all the information relating to a document as a record that makes up the various types of bibliographic repertoires. Considering the activity in time, metadata ordering refers to the order assigned to the bibliographic references and their access index, the catalog cards according to their headings, the access points to the databases of any type, and terms that make up the hierarchies of online resources. Thus, both document ordering and document metadata ordering demonstrate, simultaneously, the anteriority and the actuality of the activity.

The activity of ordering has communicative and managerial functions, the second always aimed at realizing the first. We have chosen three main functions: reading mode, location of documents and collection management, as follows.

Ordering involves production of messages, in this case, reading modes of document collection as proposals for guiding their exploitation by moving within the space, either the traditional or electronic one. These proposals for reading modes require the presentation of documents in a certain perspective, that is, according to some aspects rather than others. As guidance mode for certain routes and views on the collection, it is possible to know what there is in it, look for interesting documents and identify others whose relationship has been established in the ordering. Exhibitions and similar activities are also included in this function in their own way.

Another function of document ordering is to locate documents in a certain collection. A code or an electronic address assigned to each document singles out the same compared to the others in a collection, allowing access by performing the

relationship between the document's record in the database and its location in a particular traditional or electronic space. In a traditionally arranged collection, the code allows the document to be withdrawn and stored again according to the order proposal. In the case of electronic addresses of documents for location, an arrangement proposal itself is not in question, unless there is hierarchical structure for navigation and access. Thus, the access points on the various bibliographic repertoires refer to the referred records, which, in turn, indicate the corresponding documents.

The activity is also adopted for the management of collections with regard to a selection policy for acquisition and disposal of documents and planning of spaces for its allocation. Regarding the best use of space, they are concerned the collection growth and conditions and the most appropriate ways of storage according to the document mediums and typology, both for preservation and access purposes.

Despite the usual managerial emphasis to selection activities and user studies, all documentary activities should be properly managed, therefore, addressed specifically, but also articulated in a process of integrated management, using methods and instruments that enable a rational, economical and effective use.

Similarly to any information organization activity, ordering should be guided by institutional objectives, the characteristics of document collection and the target audience activities for which certain information can be used.

3 Concepts and document ordering methods

To explore the theme, we selected terminology about document ordering, which we systematize below, derived from the literature produced in France, depending on the degree of generalization it presents. It is also noteworthy that this is the model applied. As each language represents its own world view, and these language differences are unclear when the theoretical underpinnings of the area are not consolidated, the French terms used as reference do not have direct correspondence with the terms adopted in Brazil, despite reflecting some meaningful solutions for the Brazilian professional practice.

We identified in the book "*Mettre en oeuvre un plan de classement*", coordinated by Calenge (2010), the most current and consolidated material on the subject of document ordering, including concepts, issues and various experience reports. The following terms were taken from the glossary of this book (CALENGE, 2010, p 191-

196): ordering (*classement*), ordering chart (*cadre de classement*), ordering plan (*plan de classement*), validated shelf marks (*cotes validées*), shelf marking manual (*manuel de cotation*), shelf mark (*cote*²⁶), shelf marking (*cotation*), tagging, shelf marking (*recotation*) and correlation chart. Entries in specialized dictionaries were also used, as follows: Jouguelet (1997, p. 135-136: ordering), Boulogne (2005, p. 67: code; p. 59: shelf mark; p. 56: ordering) e Cacaly et al. (2008, p. 55: code), in addition to educational materials by Combot (2011, p 7: ordering plan; p. 29: shelf marking with symbols, colors).

Ordering (*classement*) **of documents** is defined as the arrangement process of a collection or part of a collection of documents in furniture such as shelves or archives, resulting from transactions based on a predetermined order from a certain interest criterion.

From an instrumental point of view, to perform the activity, an **ordering chart** (*cadre de classement*) may be provided, which presents the general system of document ordering in the form of an architecture of ordering that is proposed. The ordering chart would be a map or scheme of ordering document, specially developed for implementation in a given collection.

In turn, the instrument that effectively allows to operationalize the activity according to their own systematic is the **shelf list** (*plan de classement*), that we mentioned initially. It is a text that describes the organization, the ordering rules and the points of access to the document set of a collection. The plan predetermines the organization of the documentary space, which is specified by another instrument, the shelf marking manual (*manuel de cotation*), consisting of the validated shelf marks (*cotes validées*) (described below). The ordering plan should be subject to expansion and ensure the preservation of documents. It must give each document a unique place and allow to easily and quickly find a document, and store it back in place, making it possible to maintain the initial order at low cost.

The **validated shelf marks** (*cotes validées*) are those determined by an ordering plan and listed in a *manuel de cotation*. These *cotes* respond to a spatial and intellectual organization of the documentary space that is determined *a priori*, as we said. The **shelf marking manual** (*manuel de cotation*) is the shelf marking list

²⁶ The term *cote* is from Latin *quota*, which means a distinctive mark for a unit among others from the same set. Etymologically there is no equivalence for *cote* and code.

validated in a collection, which are followed by notes of use, cross references, detailing of use, etc., and may also contain an index. Thus, the term validated shelf marks (*cotes validées*) indicates that they are shelf marks (*cotes*) previously built and according to a structure or an organizing map, different from the adjusted usage of codes of bibliographical classification systems, or even the use of codes of these systems, document to document, in the absence of adaptation, as we discuss later.

A code is a brief, simple and structured character string, whose construction is based on a set of rules, with the objective of codification. The theoretical capacity of the code should match the encoding needs, considering features of univocality and stability. As for document ordering, the **shelf mark** (*cote*) is the location data of a document in a collection. The document **shelf mark** allows identifying it and ordering it on the document sets of the collection. Its elements have structured succession that provides guidance to be given to document ordering and their location. The **shelf mark** is usually represented by a set of letters or numbers, to be reproduced in the document and in all its 'images' (entrance registrations, catalogs etc.), allowing the document to be found where it was allocated. For these reasons, the codes must be mnemonic.

The *cotation* is the operation of elaborating from a **shelf mark** (*cote*) to a document pertaining to a collection, which determines the position or location of the document. It involves a strictly intellectual step on the elaboration of the code, followed by a markedly material step that takes place with the inscription of the **shelf mark** on the document record in the database, and the production of a tag and sticking it on the document. Tagging (*rondage*), in turn, is the operation that consists of fixing the label with the code on the document. The label with the code can be produced manually or by computer. In this case, we have to remember that the label can be produced directly from the code that has been entered into the registry database.

The **shelf marking** (*recotation*) occurs when one opts for a new ordering mode that leads to a change of place of the documents and the attribution of another **shelf mark** according to the new ordination framework. Similarly to the encoding process, is the intellectual code assignment step and the material step of placing the code on the outside of the document. In the recoding process, it is functional to establish a **correlation chart** on which the new adopted codes are placed in parallel in comparison with the previous existing **shelf marks**.

In any case when documents are ordered within a library or similar institutions, captions can be produced, which are fixed to the shelves, bookcases and other

furniture where documents are displayed, in order to guide the movement through space. Also to contribute to the spatial location of the documentary sets to the user and to a fruitful circulation, several other resources are used, such as the colors coupled to the **shelf marks** and, correspondingly to the captions. It should be noted that colors do not exercise the function of ordering, as is often stated in Brazil, but a signaling function. In fact, colors have no potential for displaying attributes or characteristics to the documents, for example terms and figures related to them in a hierarchical structure. Thus, it is important to strengthen the incongruity of the term 'classification with colors', such as the example of Combot (2011, p. 29) who deals, more appropriately, of **shelf marking with symbols or colors** (*cotation avec des symboles, des couleurs*). According to him, ordering documents is done according to any criterion and a mark is assigned to this criterion, which is not alphabetic or numeric, such as a color or a pictogram, which is affixed to the item via a label or the like. Combot considers the method appropriate for some libraries, such as those for children in which a color by reading level or class, for example, is quite useful. The finding as to whether the term '**shelf marking** with symbols' and the inappropriateness of the term 'classification with colors' should contribute to the intensification of studies on signaling in libraries and similar spaces.

In the case of bookstores, some private collections of rare books and fragile documents, **shelf marks** are not adopted, otherwise their original materiality is changed. Eventually, for these cases, strips of **shelf marks** are inserted in the books or other documents, without being adhered to them.

Thus, in the ordering activity, in a broad sense, forms of representation and access can be **shelf marks** on labels attached to documents, strips arranged in the documents, email addresses of documents and metadata in general as catalog cards headers, access points to data bases, categories of online resource structures.

In information organization, we consider relevant the systematic based on processes, instruments and products, which we have adopted to address the basic concepts of document ordering. Thus, the terminology presented is structured from ordering document process, which involves bibliographic classification **procedures**, **shelf marking**, labeling and shelf marking. The instruments adopted to carry out the ordering are the ordering chart, the ordering plan, the validated shelf marks (*cotes validées*), the shelf marking manual (*manuel de cotation*) and the correlation chart. The resulting products are the ordered documents.

As for the **methods** adopted in ordering, we can talk about sequential and classificatory or systematic. From sparse indications in the literature, (SALVAN, 1962, JOUGUELET, 1997, CARRIÓN GÚTIEZ, 2002, MANIEZ, 2002, CALENGE, 2010 and COMBOT, 2011), we propose a brief systematization of document ordering methods, as follows.

The primary issue is that the type of document attribute is the guiding criterion of the method. The attributes of a document, always situational, defines it as such. The sequential method can be performed in two ways. In the continuous sequential method, each document is arranged one after the other, based on some chronology of events, which requires numerical coding; in this case, the attributes in question are the order of arrival, the issues of a magazine, the date of publication. In the interspersed sequential method, each new document can be inserted among those already arranged in space, from some of its terms of identification, such as author, title, series, for which the alphabetical coding is relevant. The classification method, or systematic, in turn, implies a hierarchical organization of its units, in a unidimensional structure, through groups based on classes and subclasses. The first method - continuous sequential - provides fixed location, unlike the other two, whose intercalation leads to a relative location of the documents.

Some situations to consider, with regard to spatiality for document ordering, are as follows.

Ordering according to the principle of provenance of the documents occurs when collections are donated to libraries, implying an ordering separately from the rest of the library collection and that considers the original arrangement. This is the original arrangement of documents according to their function in the organization or in the life of the person to which they belonged, in reference to a clearly archival approach.

There is also the ordering by material attributes (size, support, cover color) or typology of documents, often also leading to a separation of the documents of a certain type or support from the rest of the collection due to the quick access requirement (the reference books allocated at the entrance of a library) or preservation (special conditions of temperature and humidity, as in the case of audiovisual, or need for fixed location for fragile documents).

Another situation concerning spaces is the information networks, such as scientific information, in which the allocation of documents takes place in every library

that makes up the network, so that the document code in the registry of databases must be preceded by the code from the library that owns the document.

Calenge (2010, p. 191-196) says that the codes can be based: on aspects such as the format or entry order, usual in private collections; on a classification system adapted to the collection, in which the list of validated shelf marks (*cotes validées*) is constructed from a pre-existing classification system; or on an ordering plan, which includes *cotes validées* built locally. We noted that, in all cases, most authors and examples of professional practice indicate that an ordering plan is concerned. However, for Calenge only in the latter case there would be an ordering plan.

Calenge (2010, p. 193) also addresses another situation in which the codes are produced from a pre-existing classification system: the choice of the code takes place in the moment of analyzing each document and according to the adopted system. For him, these codes are not called *cotes validées*. In fact, though this procedure, the adaptation of a classification system for a prior and local production of codes does not occur. Thus, the classification structure of the system is independent of what would be most appropriate to the collection of documents. This situation is usual in Brazilian libraries, especially those that take the call number to use DDC or UDC and the Cutter table of author names, according to the Anglo-American model. In turn, in France, according to Combot (2011, p. 30), for the production of codes, generally a notation of DDC is used, followed by the first three letters of the author's name. He explains that in the country, the Cutter table of author names is not taught, and questions how to distinguish authors with the same letters.

As for the French model, Calenge (2010, p. 13-15) informs that the so-called *cotes validées* are now used in the *Bibliothèque d'Public Information*, of the Centre Pompidou, in Paris, in the form of a restrictive list of UDC notations, due to the uncontrolled use of different codes that occurred, in such way that the principle of the ordering plan was originated from UDC. The motivation, according Calenge, was the search for clarity and rationality of spaces. For him, the ordering plan is the result of the activity of grouping documents by the production of codes in a structured way, according to the documentary masses at stake, the public concerned and the architectural limits. The final document represents the formalization of the physical addressing system of the documents proposed, especially for collections of free access. To Calenge, the ordering plan is a tool that makes the documentary policy produced for a specific audience visible. At the same time, it appears as a management

tool which, as such, is characterized by procedures, therefore a daily working tool of the personnel that allows each to adapt, discuss and challenge (p. 15). We observed that, as in the case of understanding, relatively consensual, the need for building thesaurus according to each system, the ordering plan that, by its own origin and definition is local, is desirable.

The distinction of models points to solutions with its own characteristics. On the one hand, the French model adopts the ordering plan as an instrument that maps and operates the document ordering, serving as an instrument for reading the collection and circulation around it, access to items, selection and disposal, preservation and optimization of space. On the other hand, the Anglo-American model has the shelflisting for internal use as an instrument for the control of items in the collection, especially one that is available directly to the public. With the computerization of the management processes of library collections, control modules of the items in the collection have been developed, enabling more powerful resource management as those that account for existing items, acquisitions, disposals etc. This module does not work, however, as an instrument for document ordering, according to the functions identified for the activity.

On the other hand, the solutions presented for the constitution of the call number are quite complete for individualization, particularly of books. It would be noteworthy to go back in time in search of understanding the construction of these proposals, as done by Weitzel (2012). From the analysis of studies by authors of the nineteenth century that addressed collection development, the author found that they were heavily guided by ideas of policy, planning, plan and criteria to develop collections and also structure to their formation. Weitzel (p. 182) cites the main current models of structure for collections known in Brazil, among them the *Conspectus*, from the *International Federation of Library Associations and Institutions* (IFLA), in which the themes and subjects of collections are determined by exemplary concentration in classes and subclasses of a classification scheme. Thus, for the consolidation of theme document ordering, both historical and contemporary proposals lack exploration.

4 Document ordering and bibliographic classification

Understanding the document ordering activity demands resuming the role of bibliographic classification. As we have seen, the activity of ordering or arranging documents is usual in French in the form *classement*. We located the use of the term

by Suzanne Briet, to whom (1951, p. 24) document ordering is not done the same way as for sales in a bookstore, for displaying in an art museum and for consultation in a specialized library. So, Briet demonstrates the breadth of the term ordering to address the activity according to different contexts. Currently, according to Jouguelet (1997), the French term *classement* is adopted to indicate the **document ordering in the space**, i.e., the material operation of putting in order, physically allocating the documents, relative to each other, while the term *classification* refers to the **classification**, as an intellectual operation for carrying it out. This distinction is recurrent in contemporary literature published in France.

In an archival approach, in France, according to Hudon (2009, p. 70, 76-77), we use the term shelf list (*plan de classement*) more often than classification plan (*plan de classification*). This instrument has an essentially hierarchical structure, comprising the list of major functions and administrative activities in an institution, or major activities of an individual, for the organization of documents corresponding to them. The goal is to preserve the context of the creation of a document and, therefore, determine its primary value (administrative, financial or legal value) and its secondary value (historical testimony value). As we can see, the preference for the term ordering plan seems emblematic of the idea of ordering as arrangement of documents according to some principle of interest: the term indicates that the ordering plan itself is not a classification one, which is, in fact, the method adopted for that. Such as the aforementioned book, produced in French, they use the term classification plan in Canada, so that is also adopted in Brazil.

In Spanish, in the classic book "*Manual de Bibliotecas*", Carrión Gutiez (2002, p. 391-402) deals with the activity of ordering collections (*ordenación de los fondos*), whose broadest sense is evident when discussing different methods of classification, as the arrival order. The author addresses the location code of the document by the term *signatura topográfica*, usual in Spanish, presenting generic definitions representing the ways to produce a code to documents listed in their records in catalogs.

In Portuguese adopted in Portugal, we identified the term *arrumação*, adopted both for document ordering on the shelf, and for the ordination of records from a catalog (FARIA; PERICÃO, 2008, p. 73). The term *cota* is defined as the symbol used to indicate the document location code on shelves, such as the call number. In turn, the authors claim that, in classification, specifically, the term *número de chamada* is set

from the classical notation made from the classification number and entry elements of the record that represents it (in general, the author's name) (p. 210 and 524).

In French, *cote* is the term adopted for the code that individualizes the document in a collection, as defined by Calenge (2009, p. 193) and Boulogne (2005, p. 67). We can say that *cote* appears as an element of specific terminology on the subject in French, as distinguished from code, used to code in the ordinary language sense. The French technical term *cote* is usual in the presentation of the results of searches in databases of France, just as the term *cota* is adopted in databases from Portugal, to indicate the location codes of the documents represented in the records.

In English, the term shelf mark historically indicates the mark or code, written or fixed in a manuscript or printed book, indicating its location in a specific library. The term was a precursor of the call number, established as a unique code printed on a label attached on the outside of an item in a library collection, also shown in the bibliographic record representing the item in the library catalog, to identify the specific copy of the work and give their relative shelf location. In general, it is used in libraries, according to the structure described above, that has as its first ordering element a number of classification (REITZ, 2013).

Both French terms presented above, relating to processes, tools and products of document ordering, as the terms adopted in other languages to distinguish document ordering and bibliographic classification, as location code and called number, demonstrated the lack of our own conceptual system and an established terminology on the subject in Brazil. However, in the book "A ordem dos livros na biblioteca: uma abordagem preliminar ao sistema de localização fixa", recently published in Brazil, the term document ordering is adopted and the authors acknowledge that there is negligence in the study and systematization of the activity, as this is a resource for access to information (PINHEIRO, 2007, p. 19). Another Brazilian differential is the literature on the underpinnings of documentary languages in which traditional systems of bibliographic classification are problematized through structural and functional aspects.

Thus, if document ordering is to propose an arrangement to them in a certain space, the key of ordering is the adopted method, which can be sequential or classificatory. The classificatory ordering, the most explored in the literature, also presents more questions to ordering as a subject of study.

Classification in a general sense involves the intellectual process by which we recognize things because they have similarities with each other, forming a set of things in relation to another. **Bibliographical classification**, in turn, is an intellectual operation that involves identifying document features, joining those that have similar characteristics and separating those with different characteristics. For grouping documents according to the classificatory method, their attributes are adopted in a collection, such as document type (book, journal, bibliography, dictionary, encyclopedia, patent), content type (genre, biography), literature origin (Spanish, English), audience (children, women), subject, others.

As we can see, document ordering does not refer solely to the subject attribute because it is not about ordering the knowledge of the documents, but to order the documents according to the various characteristics relevant to the purpose of communicating information. However, the bibliographic classification was developed as belonging to the so called Subject Representation studies. From this perspective, understanding that the bibliographic classification works only with the subject attribute was being settled in the literature, in the teaching and in the thought of the field. Classification is the logical operation of identifying the attributes of the documents as a whole throughout the collection, not only as to its thematic content.

It is worth bringing Sayers's observation (1918, chapter V) a century ago: analyzing DDC, he regarded the fact that documents do not address only the specific issues, just as encyclopedias and journals include various issues; and others, such as literary books have in their form the way issues are presented, as well as books dealing with a subject according to a certain approach, as a historical, theoretical, others. He claims that it is necessary to distinguish between form and subject, proposing a classification rule in which first, the search is done by the subject, then by the form, except in cases where the form is essential. A few years earlier, the author (SAYERS, 1915, p 43) made the following systematization for cases of using form in bibliographic classification: works composed of various subjects, works in which the form prevails over the subject and works in which specific subject are treated according to particular viewpoints.

To carry out the bibliographic classification process, the **bibliographic classification systems** that are among the documentary language, those which are characterized, in terms of structure, by a hierarchy, and in terms of its function by the vocation to document ordering.

As we mentioned, some bibliographic classification systems are more widely used. These systems have, typically, a general classification scheme by topics or subjects, auxiliary tables (that refer to common features that can be used in any class of the schema, such as shape, time, space, and viewpoint) and an alphabetic index that facilitates access to diverse perspectives under which a term can be addressed (e.g., child psychology, education, anthropology, religion) as well as the class or subclass of the most appropriate hierarchical structure to the objectives of classification.

Considering the advantages brought in the use of bibliographic classification systems for document ordering, relating to cognitive parameters that they provide for navigation between documents, to the largely widespread use of alphanumeric coding and familiarity of the public as to its structure, the exploitation of their characteristics and forms of use allows to broaden the understanding of these systems, as follows.

One of these features is the centrality of the book. Bibliographic classification systems have been developed to tackle, in particular, documents of the typology of the book, such as fictional literature, among others, and monographs, and reference works (encyclopaedias, dictionaries) and journals. The adequacy of such systems is related to the documentary typologies in libraries throughout the ages. For other typologies as audiovisual documents, reports, patents, among others, individualized ordering and coding modes are considered more appropriate. Although less common, the term documentary classification system has been adopted for a century to characterize the UDC, because it would be applied to the ordering of various types of documents, allowing the ordering of bibliographic repertories and catalogs; works in libraries; notes, observations, various documents extracts intended for study and personal work; journal summaries; graphic documents, illustrations and photographs; patents; industrial catalogs, and other documentary applications (CLASSIFICAÇÃO..., 1907). Likewise, Salvan (1962, p. 10) addresses the UDC as a documentary classification by origin, while catalographic and bibliographic, aimed at all types of documents. The pretension was that the UDC would serve any type of document classification and on any medium, as well as its metadata. On the other hand, the term 'bibliographic' does not refer, in its origin, only to the book type of documents, but this relationship is usual in the field, showing the vocation of traditional systems for books and monographs in general.

When it comes to traditional systems of bibliographic classification, the implications of the emphasis on normativity that still characterizes the mentality of the field, also come into question. The problem arises because of the rules as the preferred arena for legitimacy and understanding of procedures. Despite this prevailing thought, one should seek for the grounds from which documentary instruments origin so that they can be used productively.

Bibliographic classification was adopted for document ordering by subject, document typologies and others. Subsequently, **indexing** has also become adopted by enabling the retrieval of document subjects. However, we must address bibliographic classification and indexing as two basic and distinct operations of information organization, as follows.

As bibliographic classification is used in order to carry out the document ordering, the representation is generic because it is based on the document as a whole. The product of this type of representation is a classificatory notation based on a linguistic representation referring to a class.

In indexing, the representation is more specific because the aim is to exploit the intellectual diversity of the document. The product of the representation, the descriptors (or descriptor terms), acts as thematic access points that refer to records from the database, which indicate documents (with the exception of cadastral databases that refer to records, as we mentioned).

Unlike bibliographic classification systems, the classificatory structure of thesauri, idealized for indexing purposes, is just a reference to the global organization of the instrument. The focus on the indexing is the concept, or combination of concepts, not the class on which it resides. The position of the concept in the classificatory structure of the thesaurus causes it to be marked by the class characteristics where it is inserted, but unlike the bibliographic classification, the class is relativized because the focus shifts from the structure to the concept represented by the descriptor.

In indexing using subject headings lists, it is common for each unit to be combined with a classificatory notation, so that there is no duplication of effort between indexing and classification. The problem is that in addition to the linguistic and pragmatic constraints presented by the usual classification systems, lists as the *Library of Congress Subject Headings* or the *Sears List*, according to Lara (2002, p. 9), are instruments that do not allow effective control of the vocabulary because they do not

have a structural base of support, which worsens when their headings are translated by the institutions that use them and recorded in subject headings catalogs.

Document ordering is distinguished from the descriptive representation and thematic representation because they operate from the exploration of several possible and desirable data elements for control and search, while that has as object of representation attributes of the documentary unit in its globality. The fundamental aspects of document ordering are the materiality of documents and the spatiality on the place of the documents in a collection together with the others that compose it. Thus, document ordering refers to the treatment of documents in relation to a considered central attribute, while the descriptive and thematic representation operate respectively with the various elements of identification and thematic of the documents.

The activity of ordering in a broad sense refers to an arrangement of documents or document metadata that allows navigation and access. Although it may include a same record in various categories in the case of metadata documents, the ordering activity never offers multidimensional representation that characterizes the databases, through which several entries or access points to records of documents are possible. The activity of document ordering or document metadata ordering has, as a matrix, a hierarchical or unidimensional structure (as developed in the next section) or a sequence (numeric or alphabetic) that determines the basic arrangement. In databases, using descriptive and thematic representation, there are more possibilities of choice and crossings may be made, while from the lists ordered sequentially or hierarchically, the cognitive way to go is always linear.

The classification action is an operation adopted in all processes of information organization - cataloging, indexing, preparing summaries and document and metadata ordering (sequential, not continuous, and systematic) - as well as in subsequent processes of search on databases or ordered collections. Due to the lack of distinction about the role played by classification in each process of organizing information, the equivalence between **classification** and **information organization** is increasing. If in the mid-twentieth century, classification studies were representative of what we now understand as information organization, since it was still developing, a deepening on each of its specific processes, at present, there is a lack of recognition about the advances made. Studies on information organization, assimilated in classification studies, instill a simplification on the understanding of their processes, since these are reduced solely to the activity of identifying a core attribute of the document, obscuring

the specificity of each activity, and its function, particular methods and problems. Maniez (2002, p. 231) speaks of the use of the English term classification by Lancaster, when it comes to indexing as a classification process, exemplifying the indexing of a document by the terms 'elementary school', 'television' and 'reading skills'. For Lancaster, the example shows the assignment of this document to the three classes identified by these terms, and not in particular to the concept they represent, as we treat earlier. To Maniez, it is necessary to distinguish the specific uses of the term classification in Information Science.

5 Linguistic foundations of document ordering

The theoretical and methodological development of documentary languages, in order to support its construction and upgrade, has contributions from Linguistics and Terminology as a way of promoting linguistic communication skills according to the terminology used by the public. We can say that Gardin and others inaugurated a linguistic basis of documentary processes in the 1960s, although experiments had already been proposed before.

The usual definition for indexing languages is the instruments adopted for the representation of document content, with the purpose of retrieving them from databases. For their effective functionality, they are designed or adapted for use in every documentary context. Various types of documentary languages were constructed and used in the course of its development, such as bibliographic classification systems, lists of subject headings and thesauri. However, some intrinsic characteristics of bibliographic classification systems in the context of documentary languages, and other specific features of the most popular systems, should be highlighted.

A documentary language consists of terms and relations between them. The three types of relationships that characterize it are (CINTRA et al., 2002, p. 45-46):

- Hierarchical relationships, whose superordinate and subordinate terms can be of genre/kind or whole/part type;
- Equivalence relations, in which, to a preferred term corresponds to one or more non-preferred, with all of them acting as search key by the user; the relations perform the vocabulary control, i.e., control of natural language ambiguity; and

- Associative relationships, those that are of interest for search and are not explained previously, as cause and effect, process and instrument etc.

- From another point of view, Gardin (apud CINTRA, 2002, p. 35-36) addresses the three basic elements of documentary language:

- Lexicon: descriptors list, duly filtered and depurated;

- Paradigmatic relationships: composed by essential relationships and generally stable between descriptors, whose logical-semantic network corresponds to the organization of descriptors in a form that could be called classification; and

- Syntagmatic relationships: consisting of the contingent relationships between descriptor terms, relationships that are valid in the particular context in which they are used; the phrases are expressions constructed by syntactic rules whose coordination between terms constitutes an issue.

The paradigmatic relationships are the units that make up the documentary language as their categories (a documentary language itself), while the syntagmatic network is the use of these units in appropriate combinations performed in indexing and search (the indexing language in use).

Member of the research group coordinated by Gardin, Coyaud (1966, p. 22-23) addresses two types of relationships between terms in a documentary language:

- Analytical relationships: set of relationships explicitly indicated between the terms of the lexicon for documentary analysis; they constitute the a priori, innate part.

- Synthetic relationships: those established between the terms of the lexicon at the time of documentary analysis; private bonds they form between the terms of the lexicon, in indexing documents, are not predicted in their individuality, becoming a posteriori.

Coyaud indicates the paradigmatic organization of a documentary language as one founded on analytical relations, from what we infer that the syntagmatic organization would be one based on synthetic relations. He states that the difference between analytic and synthetic relations is not related to nature, but to use, and exemplifies the analytic relationship with the inclusion relationship, usually present in hierarchical classifications. According to him, Gardin (1964), but also Grolier (1962), attribute the analytical relationships to the plan of thought and synthetic relationships to the speech plan. Thus, the analytical relationships are those stable in a documentary language in the sense that they are its constituent elements, and not a result of the combinations carried out when using them, such as the synthetic relations. It should

be noted that, for Gardin, the constituent relations of the analytical paradigm network are always hierarchical relationships and stable only because of his study is based only on classificatory languages, as we treat below.

Gardin (1966), in a study on classifications, discard the observation of codifications and other formal aspects (decimal system, alphanumeric symbols), assuming it is through their structural features it seems reasonable to define the main types of classification. In this context, Gardin addresses analytical relationships, which he defined as each relationship linking the term to the class of which it is part in a lexical organization, in this case, a classification. He calls dimensions of a classification the nature of the analytical relationships that constitute it, proposing, therefore, two types of classifications: unidimensional and multi-dimensional. The first presents an analytical relationship, while the multi-dimensional organization is identified after the separation of the different categories of analytical relationships underlying a classification. Coyaud quotes Gardin (1962) that, in the 1966 text, discusses about univocality and multivocality of a classification referring to their dimensionality: a classification is univocal when a term appears in one place and multivocal when the same term is allocated in different classes, by the same or other analytic relationships, involving the repetition of terms.

Referring to traditional systems of bibliographic classification, however, Gardin (1966) proposes to distinguish between a real unidimensionality and an apparent unidimensionality. The real unidimensionality is that, for example, of taxonomies, typical of the natural sciences in which the constitutive relationship is effectively unique at all levels and for all classification groups. However, uniqueness of the dimension may be apparent. He states that, under the qualification of hierarchical, often attributed to this unique relationship, different types of relationship are identified, such as qualification, location, instrumentality etc., which have nothing in common with a relationship of inclusion. For Gardin, while the taxonomic systems present a real unidimensionality, in general bibliographic classification systems trigger an apparent unidimensionality, except in some of its parts, as in the nomenclature of natural species, chemical bodies etc. Thus, other kinds of relationships are anonymous and, therefore, are confused with the strict hierarchical relationship. In turn, Coyaud (1966, p. 35-376) exploits the idea showing the not explicit diversity of analytical relationships of the UDC, in which several relationships are amalgamated into a single inclusion relation.

Later, Lara (2002) analyzed the DDC and the UDC and made a comparison between the two systems, reinforcing the analysis of Gardin and Coyaud. She states that both have restrictions for including, in a single hierarchy, different types of relationships between units. Hierarchical relationships can be observed according to the nature of the relationship (with the gender/kind being the only genuine hierarchical relationship), or according to the form of presentation (when both the relationship gender/kind and the whole/part are presented in the form of a tree). Lara (p. 6-7) claims that the world is not organized in a hierarchical manner (i.e., in a unidimensional mode) and various types of relationships can occur between concepts such as relationships named associative nonhierarchical, the cause-effect type, producer-product, action-object, agent-instrument, activity-agent etc. For her (p. 4), systems such as the DDC and the UDC are more apt to provide a basis for the formation of stocks and for the reproduction of knowledge to promote the flow of information.

As for the unidimensional characteristic of hierarchical systems explicit by Gardin, it is important to distinguish the restrictions inherent to bibliographic classification systems and those specific to the traditional systems of bibliographic classification, as the UDC and the DDC. Bibliographic classification systems are inherently hierarchical systems, therefore, unidimensional. The attempt to contemplate other dimensions, by inserting relationships between several terms, de-characterize these systems, turning them into apparent dimensional, therefore, impaired in its consistency and functionality. Bibliographic classification systems, as well as the lists of subject headings, are not considered documentary languages, such as the thesauri. What is at stake is the relationship between the vocation of each documentary language, the way it is built and what it is used for. As an example, we have the first proposals for corporative taxonomies which, quickly, incorporated other types of relationships between the terms beyond the hierarchical relationships, aiming at functionality as an information retrieval instrument.

In this regard, the contribution of Gardin is, among others, the discovery that the choice of relations between the constituent terms of an indexing language expresses the originality of its structural point of view and, therefore, its place in the typology of the languages in general. Gardin (1966) advances in the identification of multidimensional classification typologies, addressing an empirical distinction between two categories of analytical relationships:

- Semantic: refer to the order of the terms (i.e., the notions they designate) according to the definitions currently accepted by a community and considered inherent characteristics of the entities, i.e., always and in every circumstance valid; the classification may be regarded as the image of that order (temporary, local), in which the primitive meaning of the terms determines its place in the overall organization; these are essential relationships.

- Syntactic: they do not refer to the essence of the entities, but its particular function in a particular field of view; properties considered to group the terms are contingent, i.e., they are valid only in a restricted context, one in which the classification is developed; they are functional relationships.

But Gardin notes that the documentary classifications are based on both essential and functional relations, configuring mixed classifications, or analytical-synthetic, which vary, from a strictly semantic order of the purest taxonomies to the exclusively syntactic order, such as faceted classifications. In the latter case, a term may appear anywhere on the classification in the different functional classes to which it belongs. For example, in a relative classification regarding the pharmaceutical industry, the term iodine can compose classes that relate to raw material, final product, to reagent, etc., causing the term to repeat in the structure. The repetition of the words from different classes, due to their different dimensions is typical of faceted classification systems. The hierarchical structure includes relationships of another type, namely, the syntactic level because, in the classes A, B and C, relationships are established because it identifies that there are terms that are members of A, members of B and members of C. Therefore, there is a multivocal classification, as the terms are repeated at various locations.

The idea of facet, proposed by Ranganathan, innovated by isolating the characteristics of terms and arranging them in the structure according to the views of interest. The theory of faceted classification, as it became known, was one of the bases for the development of documentary languages geared for the retrieval of information, such as thesauri.

The approach of classifications to a proper language, as Gardin developed, enabled an understanding of the paradigmatic axis and the syntagmatic axis of documentary languages. This theoretical framework of language allowed to distinguish pre-coordinated documentary languages, such as faceted classifications and others, in which the terms are combined in advance, making the system, or are minimally

combined at the time of document representation, and post-coordinated documentary languages, as thesauri, in which the combination of terms occurs in document representation and search in the system. It is also possible to observe the distinction between documentary languages of classification and documentary indexing languages, the first focused on the document as a whole, in a linear and unidimensional organization, and the second for its thematic content, on a multidimensional perspective.

The linguistic understanding of information organization does not happen out of a pragmatic perspective, so that the production and use of traditional systems of bibliographic classification should focus on features of document collections and specifics of the target audience. Faced with the task of providing a scientific character to some conceptions of common sense, we identified three issues, which we comment below.

The first refers to the usual characterization of traditional systems of bibliographic classification as **encyclopedic**, although they concern, in fact, the universalizing claim to encompass, in a single system, the whole universe of knowledge (LARA, 2001). The universalization, however, does not occur because, according to this author, in a semiotic approach, what is registered is always partial faced with the infinite possibility of encyclopedic interpretation. Lara (1999) has focused on the concept of encyclopedia demonstrating not only its complexity, but its relevance to an understanding of the area, as opposed to the dictionary idea. While the dictionary presents the lexicon of a language and their definitions in an alphabetical presentation, the encyclopedia is made by the notion of chaining, providing the reader related entries that allow various reading paths. Later, Olga Pombo explored the concept of encyclopedia as one that provides to the area the categories of thought that explain and support (POMBO, 2010).

A second issue relates to the fact that the **disciplinary matrix** present in traditional systems of bibliographic classification is not representative of the contemporary world, and complicates its use and update. So, if forcing a hierarchical system to be multidimensional seems not feasible, trying to encompass in it all knowledge relating to a period of very ancient time makes evident a homogeneous and stable worldview and, therefore, anachronistic. The third issue is that, in the library science culture, information was continually taken as data, i.e., the information would exist a priori and the representation would be the effort to rescue it from the document.

This perspective led to the idea that in the representation there is a trust relationship between the content of a document and the expression used to 'represent it'. As a result, based on Lara (2001), the classification product would be a **reproduction**.

As for the three issues, we observe, respectively, that the idea of supposedly universal system persists. These systems, adopted in the absence of perception of meaning of historical dating of a particular approach to knowledge, and departing from fixed points of articulation to be mechanically equivalent to the contents of documents, whose interpretation is also unique.

To Lara (2001), information is always a construction, because the generalizations made in the documentary activity are a cultural product, the result of the experience with the reality, therefore, they are neither neutral nor have a purpose in themselves. The establishment of information systems is an operation charged with intentionality that pursues specific goals. The pragmatic needs determine the cutouts, namely, the segmentation of the contents, in such a way to enable effective information flow. Consequently, even a documentary universe can be organized and represented in various ways, and based on the different purposes for which they are intended.

If language is the main expression of different forms of cultural organization, it is justified to search in the theories of language support for the construction of the documentary language, as well as references to achieve forms of more adhesive mediation to information systems (MOREIRA; LARA, 2012).

Thus, under the linguistic point of view, the classificatory ordering is based on a unidimensional structure, which has its own characteristics that allow us to assess accurately the various bibliographic classification systems. The implications of its apparent unidimensionality pass by the fact that these systems are not, in fact, neither unidimensional, having only one type of analytical relationship, nor multidimensional, presenting multivocality from consistent, organized repetitions. It is necessary to distinguish essential and contingency aspects, the first always necessary for the perception of the latter. Gardin's analysis allows to identify that the problems of the traditional systems of bibliographic classification are not intrinsic to this kind of system. The fact that some of them are not effectively unidimensional is a contingency.

6 Conclusion

Processes and instruments based on sequences and/or hierarchies are relevant to document ordering and document metadata ordering as it is, in all cases, about allocation in space. For the activity of document representation aiming at content retrieval, multidimensionality, characteristic of thesauri adopted for the operation of indexing, should be the reference.

There is no single document ordering mode, although it is interesting to share solutions. Production arrangements that consider users, documents and institutional goals are necessary. As in all other documentary operations, the orientation is always pragmatic, i.e., based on their possible usefulness. The challenge is the refinement of methodologies based on parameters that allow the development of relevant categories to the activity.

The choice of categories is always local and provisional. Paradoxically, the documentary work requires relatively stable representation structures, due to the high cost of recoding documents and the requirement of understanding the new ways of ordering by users.

In traditional or electronic environments, none of the operations described for activity of ordering is mechanical in fact or of lesser intellectual requirement. Document ordering should play a role reviewed in the context of all the mediation actions carried out in the various information services. Developing processes of information organization provides solid base of support to operate with the changes that occur over time. In the absence of sufficiently problem-solving perspectives on bibliographic ordering of documents in teaching and research, we carry out two observations. The first, on the apparent contradiction between the emphasis on users and the unquestioned teaching of a single model as solution for document ordering; in fact, users' studies do not always consider the mediating perspective of information organization processes that allows the bond with an audience. The second observation is that there is production of a substantiating and operationalizing literature on the document ordering process, from secular studies on bibliographies and libraries, to the linguistic insights of classification instruments proposed by Gardin in the middle of the last century and its developments, as well as the contemporary French proposal, whose design and applications deserve to be known.

References

- BOULOGNE, Arlette (Coord.) (2005). *Vocabulaire de la documentation*. Paris: ADBS Éditions. (Sciences et techniques de l'Information). 334 p. Também disponível no endereço: <http://www.adbs.fr/vocabulaire-de-la-documentation-41820.htm?RH=OUTILS_VOC&RF=OUTILS_VOC>. Acesso em: 1 ago. 2015.
- BRIET, Suzanne (1951). *Qu'est-ce que la documentation?* Paris: Édit - Éditions Documentaires Industrielles et Techniques. 48 p. Disponível em: <<http://martinetl.free.fr/suzannebriet/questcequeladocumentation/briet.pdf>>. Acesso em: 1 ago. 2015.
- CACALY, Serge et al (2008). *Dictionnaire de l'information*. 3. ed. Paris: Armand Colin. 296 p.
- CALENGE, Bertrand (Coord.) (2010). *Mettre en oeuvre un plan de classement*. Villeurbanne: Presses de l'ENSSIB. (La Boîte à outils, 18). 199 p. p. 191-196: Glossário. Publicado originalmente em 2009.
- CARRIÓN GÚTIEZ, Manuel (reimpr. 2002). *Manual de bibliotecas*. 2. ed. de 1993. Madrid: Fundación Germán Sánchez Ruipérez. 766 p. p. 391-402.
- CINTRA, Anna Maria Marques et al. (2002). *Para entender as linguagens documentárias*. 2. ed. São Paulo: Polis.
- CLASSIFICAÇÃO bibliográfica decimal (1907). *Bulletin de l'Institut International de Bibliographie*, Bruxelles, ano 12, n. 68-70, não paginado.
- COMBOT, Philippe (2011). *Initiations aux techniques documentaires*. Disponível em: <<http://www.youscribe.com/catalogue/tous/education/cours/initiation-aux-techniques-documentaires-385019>> e em <<http://combot.univ-tln.fr/lea/index.html>>. Acesso em: 1 ago. 2015.
- COYAUD, Maurice (1966). *Introduction a l'étude des langues documentaires*. Paris: C. Klincksieck. 148 p. (T. A. Documents).
- FARIA, Maria Isabel Ribeiro de; PERICÃO, Maria da Graça (2008). *Dicionário do livro: da escrita ao livro eletrônico*. São Paulo: EDUSP. 768 p.
- GARDIN, Jean-Claude (1966). Elements d'un modele pour la description des lexiques documentaires. *Bulletin des Bibliothèques de France*, n. 5, p. 171-182. Disponível em: <<http://bbf.enssib.fr/consulter/bbf-1966-05-0171-001>>. Acesso em: 1 ago. 2015.
- HUDON, Michèle (2009). Le traitement du document. In: SALAÜN, Jean-Michel; ARSENAULT, Clément (Coords.). *Introduction aux sciences de l'information*. Montréal: Les Presses de la Université de Montréal. (Paramètres). 235 p. p. 53-100.
- JOUGUELET, Suzanne (1997). Classement. In: CACALY, Serge (Coord.). *Dictionnaire encyclopédique de l'information et de la documentation*. Paris : Nathan. 635 p. p. 135-136.
- LARA, Marilda Lopes Ginez de (1999). *Representação e linguagens documentárias: bases teórico-metodológicas*. São Paulo. Tese (Doutorado) – Escola de Comunicações e Artes/USP.
- LARA, Marilda Lopes Ginez de (2001). O unicórnio (o rinoceronte, o ornitorrinco...), a análise documentária e a linguagem documentária. *DataGramaZero: Revista de Ciência da Informação*, Rio de Janeiro, v. 2, n. 6.

LARA, Marilda Lopes Ginez de (2002). *Dos sistemas de classificação bibliográfica às search engines (I)*. São Paulo: Associação Paulista de Bibliotecários. (Ensaio APB, 90).

MANIEZ, Jacques (2002). *Actualité des langages documentaires: les fondements théoriques de la recherche d'information*. Paris: ADBS. (Collection Sciences de l'Information, Série Recherche e documents).

MENESES, Ulpiano T. Bezerra de (1994). Do teatro da memória ao laboratório da história: a exposição museológica e o conhecimento histórico. *Anais do Museu Paulista*, v. 2, n. 1, p. 9-42. Disponível em: <<http://www.scielo.br/pdf/anaismp/v2n1/a02v2n1.pdf>>. Acesso em: 1 ago. 2015.

MEYRIAT, Jean (1981). Document, documentation, documentologie. *Schéma et Schématisation*, 2º trimestre, n. 14, p. 51-63.

MOREIRA, Walter; LARA, Marilda Lopes Ginez de (2012). Ontologias, categorias e interoperabilidade semântica. *DataGramaZero*, v. 13, n. 4.

PERROTTI, Edmir; PIERUCCINI, Ivete (2007). Infoeducação: saberes e fazeres da contemporaneidade. In: LARA, M. L. G. de; FUJINO, Asa; NORONHA, Daisy Pires (Orgs.). *Informação e contemporaneidade: perspectivas*. Recife: Néctar. p. 47-98. Disponível em: <<http://www3.eca.usp.br/sites/default/files/form/ata/pos/ppgci/publicacoes%20-%20informacaoContemporaniedade%281%29.pdf>> . Acesso em: 1 ago. 2015.

PINHEIRO, Ana Virginia Teixeira da Paz (2007). *A ordem dos livros na biblioteca: uma abordagem preliminar ao sistema de localização fixa*. Rio de Janeiro: Interciência, Niterói: Intertexto.

POMBO, Olga (2010). Dispersão e unidade: para uma poética da simpatia. In: LARA, Marilda Lopes Ginez de; SMIT, Johanna Wilhelmina (Orgs.). *Temas de pesquisa em Ciência da Informação no Brasil*. São Paulo: Escola de Comunicações e Artes/USP. p. 29-46. Disponível em: <<http://www3.eca.usp.br/sites/default/files/form/ata/pos/ppgci/publicacoes%20-%20temasdepesquisas.pdf>>. Acesso em: 1 ago. 2015.

REITZ, Joan (2013). *ODLIS - Online Dictionary for Library and Information Science*. Disponível em: <<http://www.abc-clio.com/ODLIS/about.aspx>>. Acesso em: 1 ago. 2015.

SAGREDO FERNÁNDEZ, Felix; IZQUIERDO ARROYO, José María (1983). *Concepción lógico-lingüística de la Documentación*. Madrid: IBERCOM. 440 p.

SALVAN, Paule (1962). *Les classifications*. 2. ed. Paris: Bibliothèque National. Preparação para o Diploma Superior de Bibliotecário. Disponível em: <<http://www.enssib.fr/bibliotheque-numerique/documents/48774-les-classifications.pdf>>. Acesso em: 1 ago. 2015.

SAYERS, W. C. Berwick (1915). *Canons of classification applied to "The subject" "The expansive", "The decimal" and "The Library of Congress" classifications: a study in bibliographical classification method*. London: Grafton. 173 p.

SAYERS, W. C. Berwick (1918). *An introduction to library classification: with readings, questions and examination papers*. London: Grafton & Co. ; New York City: The H. W. Wilson Co. 172 p.

WEITZEL, Simone da Rocha (2012). Desenvolvimento de coleções: origem dos fundamentos contemporâneos. *TransInformação*, Campinas, v. 24, n. 3, p. 179-190.

ISKO-Brazil and the strengthening of research exchange and scientific visibility in knowledge organization²⁷



Vera Dodebei

Federal University of the State of Rio de Janeiro
dodebei@gmail.com

The field of knowledge organization and representation (KO), according to Dodebei and Guimarães (2013), has become a challenge as the data bases become virtualized and information is posted on the network in multiple formats and multiple tags. If this guarantees cultural diversity in the use of resources, on the other hand, the high rate of thematic dispersion of information resources will force knowledge organization researchers to draw, together with professionals from related fields, systems and programs that help shared indexing and classification processes. Collaboration between knowledge producers and users in the contemporary world seems to be the best strategy to carry out highly costly tasks as those typical of KO area, for example bibliographic classification, indexing and documentary languages construction.

KO is a field that demands constant reflection on the creation of "tailor-made" concept systems to organize analogical and digital collections, websites, repositories, academic and commercial database and, in this sense, the experience exchange between researchers from international research centers is fundamental to the expansion of theoretical-epistemological foundation of scientific and technical research.

The Brazilian Society for Knowledge Organization (ISKO - Brazil), created on October 30th, 2007, is the Brazilian chapter of the "Internationale Gesellschaft für Wissensorganisation", or International Society for Knowledge Organization (ISKO)

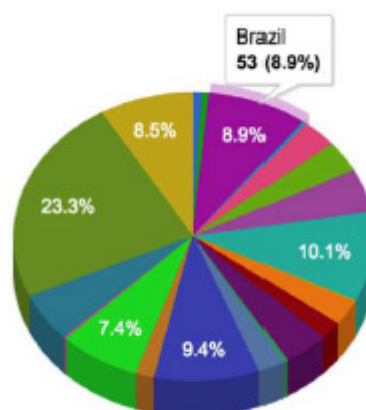
²⁷Round Table: Visibilidade científica e interlocução internacional em organização do conhecimento. III ISKO-Brasil. Marília, SP. 2015.

<http://www.isko.org/index.php>, registered in Frankfurt, Germany, on July 22nd, 1989. ISKO-Brazil is a nationwide association, with legal headquarters in Marilia, SP. Its administration is linked to the institution where the president is affiliated, for the 2013-2105 term, in the state of Rio de Janeiro.

ISKO-Brazil aims to promote research, education, development and application of methods for knowledge organization (KO) in general and in specific fields in the following ways: promoting, at a national and international level, the exchange of information and experiences among researchers and practitioners in the KO area, conducting scientific meetings and events in the area, organizing and promoting research of visibility in publications in the field, working with the International Society for Knowledge Organization in achieving its objectives and promoting scientific cooperation mechanisms, as well as international promotion and insertion of studies and research carried out by its members.

Our objectives, in the short term, have been strengthening institutional bonds between the two associations. Such a bond will be manifested in the coming years with the implementation of the following goals and actions: 1) to organize two national events (2015 and 2017), in which activities with international keynote speakers are planned, such as lectures, short courses, discussions with research groups. In these events, consolidation meetings of research groups already in development and encouraging the formation of new groups are planned, adding Brazilian and foreign researchers; 2) to organize an international event in 2016, in which short courses with international experts, and two magna conferences: opening and closing are planned; and 3) to publish the electronic annals with the papers presented in national events, which boosts cooperation instruments, through the encouragement of co-authored publications, intensifying partnerships in progress and encouraging new partnerships.

In its nearly twenty years of existence, International ISKO has expanded considerably in all continents, and today its team of associate representatives come from the following countries: Algeria, Benin, Brazil, Cameroon, Canada, China, Spain, United States, France, Ghana, India, England, Iran, Italy, Morocco, Nigeria, Poland, Singapore, Taiwan and Tunisia. The latest information indicates that its membership includes 596 researchers, including highly experienced researchers and others, fewer in number, researchers in training. It must be taken into consideration that 18.3% of this total are from developing countries. It is also important to emphasize the role of Brazil in the big picture:



Graphic 1 – ISKO Associates

Source: ISKO. Available at: <http://www.isko.org/index.php>. Access on: 03.08.2015.

In less than ten years of creation, our associates board represents nearly 9% of International Association members, surpassed only by three countries - England, Germany and Poland, which indicates a strong presence of scientific research in this area in our country. Moreover, the concern with the formation of new researchers is valid. Regarding the number of partners, Brazil has approximately 20% of members who are pursuing their education in master's or doctorate (according to the associate board available on the association's website), which indicates a revitalization of the research area and the certainty that the investigative field remains productive.

The Brazilian chapter has promoted, since its creation in 2007, the organization of some events, in the years not coinciding with the international event. In 2009, because it was still a new association, it held a small event inserted in Work Group 2 at National Meeting of Information Science Research (ENANCIB). In this event, given the impact and the wide scope of participation, the Executive Board viewed the possibility of raising more consistent debate and the members decided to organize the First National event in 2011 in Brasilia. In 2013, the II National Congress edition took place in Rio de Janeiro. In this third edition the National Congress takes place in Marilia and the 14th edition of the international conference will be in Brazil in the city of Rio de Janeiro.

From the research point of view and, because the Association is configured by individual insertion of its members, there is strong involvement of various research institutions spread throughout the world to which its members are affiliated. To give an example, we present the various institutions involved in Brazil, from ISKO's

membership: Universidade Federal de Minas Gerais, Universidade Estadual de Londrina, Universidade de São Paulo, Universidade Federal do Estado do Rio de Janeiro, Universidade Federal de Pernambuco, Universidade de Brasília, EMBRAPA, INMETRO, Universidade Federal do Paraná, IBICT, Universidade Estadual Paulista, Fundação Getulio Vargas, among others.

One of the indicators we may use to measure the impact caused by the cooperation between the International Association and the Brazilian chapter is the reflection on the number of associates after each edition of the scientific events. In addition to this, we have noticed the increasingly common presence of Brazilian researchers in interinstitutional exchange programs represented by Minter and Dinter (Capes) and visiting professors, although the number of published articles in the KO international journal is still low. As two indicators, we can see the presence of co-authorships of Brazilian researchers and co-authorships among national and international researchers and also the establishment of the correlation index among researchers and their institutions in order to list the existing networks - and new networks that may be created - in the KO area.

We aim to improve the quality of our events, by increasing the opportunity of personal contact with a greater number of research groups for future exchange opportunities as we can augment the number of foreign lecturers invited to come to Brazil; print the publication of the series "Advanced studies in knowledge organization and representation" in addition to the open archive, in order to intensify the socialization of research results and the opportunity to exchange new information and communication technologies already developed and in development in Brazil.

This third congress, ISKO-Brazil elected to discuss challenges and perspectives that emerge for knowledge organization in the face of cultural diversity, the state of the art of Brazilian research and scientific dialogue in Brazilian and international academic environment. The selected papers are already published in IPUB and are available on the ISKO-Brazil website, composing the series:

Estudos avançados em organização e representação do conhecimento: diversidade cultural (Advanced studies in knowledge organization and representation: cultural diversity) v.3, 2015, available online at the society's website. The 14th International ISKO Conference in Rio de Janeiro, from 12 to 15 of September 2016 will present numerous challenges, as it will be a major event and will be open to

submission of papers for ISKO members that today brings together 21 chapters and 596 researchers.

The theme of the congress - *Knowledge organization for a sustainable world: challenges and perspectives for cultural, scientific and technological sharing in a connected society* - places the KO field, previously restricted to libraries, in connection with massive and connected ICT, as well as with research on digital humanities, social memory and economic sustainability, pointing to new cross-cutting research themes, such as information ecology, which will have visibility with the publication of print proceedings and open access in English language, which will compose the volume 15 of *Advances in Knowledge Organization*, to be published in 2016 by Ergon Verlag publisher, Germany. The planning and execution of the IV Brazilian Congress on *Knowledge Organization and Representation* to take place in 2017, places Isko-Brazil in the calendar of events supported by sponsoring agencies and reaffirms the event as a well-qualified one within the Brazilian assessment of graduate courses.

We do not have yet a study to verify for the visibility of foreign inter-institutional cooperation regarding fellowship programs for student mobility and teacher training courses abroad and in Brazil, and we need to reinforce knowledge organization fellowship programs between study groups and international research laboratories whose subjects can enhance research in the field of knowledge organization.

References

The references was made following the ABNT rules.

DODEBEI, Vera; GUIMARÃES, J. A. C. Prefácio. In: *Complexidade e organização do conhecimento: desafios de nosso século*. Rio de Janeiro: ISKO-Brasil; Marília: FUNDEPE, 2013. (e-Book ISBN: 978-85-98176-51-2)

ISKO.<http://www.isko.org/index.php>

ISKO-Brasil. <http://isko-brasil.org.br/>